4-2014

REDOX REGULATION OF PROTEIN TRANSLATION IN EUKARYOTES

Maxim Gerashchenko
University of Nebraska-Lincoln, germaximus@gmail.com

REDOX REGULATION OF PROTEIN TRANSLATION IN EUKARYOTES

by

Maxim Gerashchenko

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Biochemistry

Under the Supervision of Professors Dmitri E. Fomenko and Vadim N. Gladyshev

Lincoln, Nebraska

April, 2014

REDOX REGULATION OF PROTEIN TRANSLATION IN EUKARYOTES

Maxim Gerashchenko, PhD.

University of Nebraska, 2014

Advisors: Vadim N. Gladyshev and Dmitri E. Fomenko

Gene expression may be controlled at multiple levels, e.g., through genomic architecture, transcription and translation. In the current work, we focused on regulation of protein synthesis. Historically, the investigation of the regulation of gene expression at the level of translation lagged behind the transcriptional control because of the lack of accessible high-throughput methods. Our research has begun with the finding of the use of alternative non-AUG start codon in thioredoxin-glutathione reductase (TGR), a selenoprotein involved in redox control during male reproduction. The use of this codon, CUG, relies on the Kozak consensus sequence and ribosomal scanning mechanism. However, the CUG serves as an inefficient start codon that allows downstream in-frame initiation, generating two isoforms of the enzyme *in vivo* and *in vitro* from the same mRNA. These findings were extended with the use of systemic, proteome-wide approaches, that supported targeted discovery of initiation start sites. For this purpose, a new technology, ribosomal profiling, was employed. It embraced high-throughput sequencing and offered analyses of ribosome occupancy along the mRNA at a single

nucleotide resolution. We applied this technique to examine the interplay between transcription and translation under conditions of hydrogen peroxide treatment in *Saccharomyces cerevisiae*. Oxidative stress elicited by hydrogen peroxide led to a massive and rapid increase in ribosome occupancy of short upstream open reading frames (uORFs), including those with non-AUG translational starts, and N-terminal regions of ORFs that preceded the transcriptional response. In addition, this treatment induced the synthesis of N-terminally extended proteins and elevated stop codon read-through and frameshift events. It also increased ribosome occupancy at the beginning of ORFs and potentially duration of the elongation step. We identified proteins whose synthesis was rapidly regulated by hydrogen peroxide post-transcriptionally; however, for the majority of genes increased protein synthesis followed transcriptional regulation. Nevertheless, a number of proteins were regulated post-transcriptionally even at the 5 min time point. These data defined the landscape of genome-wide regulation of translation in response to hydrogen peroxide and suggested that "potentiation" (co-regulation of the transcript level and translation) is a feature of oxidative stress. Finally, we expanded this research to better define conditions for ribosome profiling, which are broadly applicable for studies on regulation of translation.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Vadim Gladyshev, for giving me the opportunity to work in his lab, as well as for his support and continuous guidance during my years as a graduate student. A special thanks goes to Dr. Alexey Lobanov for his guidance on bioinformatics tools.

I would like to thank my supervisory committee members, Dr. Audrey Atkin, Dr. Mark Wilson, Dr. Dmitri Fomenko, and Dr. Jaekwon Lee, for their support, assistance and suggestions. I would like to thank all my friends and colleagues for advice and assistance.

TABLE OF CONTENTS

ABBREVIATIONS

3' UTR	three prime untranslated region
5' UTR	five prime untranslated region
GFP	green fluorescent protein
GR	glutathione reductase
Grx	glutaredoxin
IRES	internal ribosome entry site
miRNA	micro ribonucleic acid
mRNA	messenger ribonucleic acid
mRNA-seq	mRNA sequencing
NGS	next generation sequencing
ORF	open reading frame
Ribo-seq	sequencing of ribosome protected footprints
Rpkm	reads per kilobase per million
rRNA	ribosomal ribonucleic acid
SECIS	selenocysteine insertion sequence
SGD	Saccharomyces genome database
TE	translation efficiency
TGR	thioredoxin glutathione reductase
TR	thioredoxin reductase
tRNA	transport ribonucleic acid
Trx	thioredoxin
uORF	upstream open reading frame
nt	nucleotide/s

CHAPTER 1

INTRODUCTION

1.1 Translation

Genetic information contained in a form of linear nucleotide sequences has to be interpreted to generate proteins. The central dogma in biology states that DNA serves as a matrix for RNA and the latter bears information to be converted into proteins. Not all RNAs encode proteins; those that do are named mRNA (messenger). Transcription is a process of synthesizing RNA whereas translation is synthesizing protein. Unlike transcription, translation represents a bigger challenge. With some exceptions, all proteins consist of 20 amino acids and there is no complementarity between nucleic bases and amino acids. Therefore an adaptor is needed to decipher the nucleic acid sequence. This function is performed by tRNAs (transfer). They are charged with amino acids by aminoacyl-tRNA synthetases and get transported into a ribosome, which is an immense molecular machine designed for peptide synthesis.

The translation machinery decodes only the portion of each mRNA called the open reading frame (ORF). It is flanked with 5' and 3' untranslated regions (UTR). There are many features located upstream, downstream and even inside of ORFs that affect translation efficiency or are hypothesized to. This doctoral thesis is focused on mechanisms of translational regulation. There are two related topics presented here: alternative translation site in a selenoprotein and a high-throughput screening method quantifying translation efficiency (TE) of multiple genes at once. These topics demonstrate the rapid pace of method in biochemistry that allowed making a move from

analysis of one gene at a time to a genome scale survey with comparable efforts spent on each study.

The purpose of this introduction is to give a brief overview of post-transriptional control and regulation in eukaryotes. It only mentions the topics relevant to the experimental scope of this dissertation and does not pretend to be a comprehensive in-depth review on the subject.

1.2 Ribosome

The understanding of ribosomal structure and function is central to the second part of the thesis. Ribosomes are the largest macromolecular complex in living cells (Fig. 1.1). In eukaryotes, it consists of one small and one large subunits, comprising four rRNAs and 79 core proteins, conserved from yeast to human [1, 2]. The large subunit consists of 46 proteins and 3 rRNA molecules (25S, 5.8S, 5S). The small subunit has 33 proteins and 1 rRNA (18S). Most of the ribosomal proteins are incorporated into ribosome at an equimolar ratio. Several recent proteomic studies of yeast cells identified nearly one hundred translation machinery-associated proteins, which are sub-stoichiometrically bound to ribosome and, in theory, could modify its translation properties [3]. Ribosomal RNAs serve both as a structural scaffold for proteins and a catalytic center. High-resolution crystal structures of the ribosome revealed that the catalytic center is almost exclusively formed by rRNA with minimal interference of proteins. Therefore, the prevailing view among researchers suggests that the contemporary ribosome evolved from a primitive protein synthesizing machine, composed entirely from RNA. This sets it aside from other cellular molecular machines, which rely on catalysis performed by proteins.

Altogether, these and other studies picture a highly complex system with a great potential for regulation.

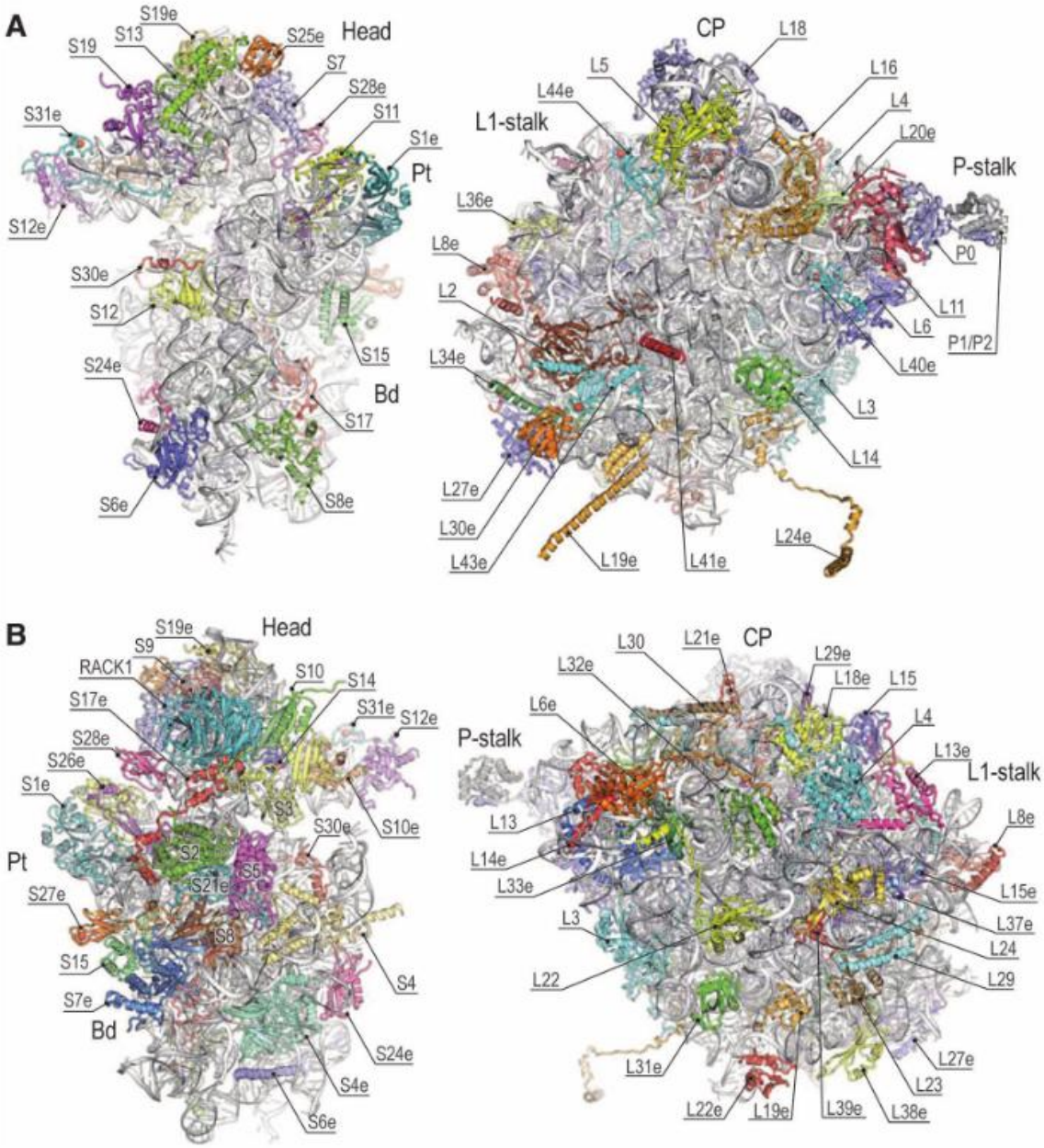

Figure 1.1. Crystal structure of the 80S eukaryotic ribosome. (**A**) Front view of the large subunit (left) and small subunit (right). (**B**) Back side from the same subunits. Images are taken from [2].

1.3 Structure of the catalytic center of the ribosome

The key catalytic event carried out by a ribosome is the peptidyl transferase reaction, which results in an addition of an amino acid to the C terminus of a nascent peptide. The ribosome contains three sites capable of tRNA binding, called A, P and E. They are located next to each other in a catalytic core of the ribosome. Each of these sites is formed at the interface between the large and the small ribosomal subunits and together they are organized into a channel that passes through the ribosome. During the translation event, the mRNA transcript is stretched inside of this channel and decoding occurs by pulling the transcript 3 nucleotides at a time. An additional tunnel is used by a nascent peptide chain to get out of the ribosome.

Translation of a peptide can be split into three phases, called initiation, elongation and termination. Each step requires numerous protein factors and small molecules assisting the ribosome and is fairly complex and well orchestrated [4, 5].

1.4 Selenocysteine as a unique amino acid

The pool of amino acids used to compose proteins is usually limited to 20 different types in eukaryotes. However, several proteins, including 25 proteins in mammals, utilize selenocysteine, a very rare amino acid [6]. Not just it is rare but it also requires an additional protein network to incorporate it into polypeptide chains. Although selenocysteine is used by all three kingdoms of life, some species do not use it at all [7, 8]. Budding yeast are one of these exceptions. Another feature of selenocysteine is that it is encoded by the UGA codon, which is generally decoded as the stop signal, terminating protein synthesis. Therefore, there is a need to distinguish selenocysteine-encoding UGA and terminator UGA, which is achieved by using a special tRNA and a SECIS element

(selenocysteine insertion sequence) within the mRNA sequence. The details of selenocysteine incorporation are different in bacteria, archaea and eukaryotes, but all of them share these similar components.

1.5. Cap-dependent and cap-independent translation

Eukaryotic cell is populated by many kinds of RNA transcripts. In order to distinguish coding mRNA from non-coding nucleotide sequences, translation machinery relies on a secondary modification at the 5' end of mRNA called "cap". At the same time, the 3' end has to be polyadenylated. Most of the mRNAs in a eukaryotic cell are capped and polyadenylated. More than a dozen of translation initiation factors secure the recognition of these signals while loading the ribosome on the mRNA. Therefore, the absolute majority of genes are translated in the cap-dependent manner [9]. A subset of genes is able to undergo translation even in the absence of the cap. This strategy is broadly implemented by viruses to hijack host translation machinery and repress synthesis of host endogenous proteins. These genes attenuate the ribosome directly to the vicinity of start codon, therefore avoiding binding the cap and subsequent scanning by the small ribosomal subunit [9]. In order to do so, the ribosome binds to the IRES (internal ribosome entry site) – a region of mRNA with rich secondary structure. IRESs of different viruses and hosts have little similarity to each other and operate in various ways.

1.6 Selection of a start codon in open reading frame

In eukaryotes, the small ribosomal subunit forms a 43S preinitiation complex, containing the initiator methionyl tRNA, initiation factor 2 (eIF2) and GTP. This assembly is aided by eIFs 1, 1A, 3 and 5 [10]. The preinitiation complex is recruited to the cap structure at the 5' end of the mRNA. This recruitment also requires multiple

protein factors, such as eIF4A, E, G and poly-(A) binding protein PABP. Upon binding the mRNA, the 43S complex scans for the AUG start codon, moving from 5' to 3' direction [10]. Recognition of AUG is achieved by a perfect matching with the initiator`s tRNA anticodon. When it happens, GTP gets hydrolyzed and eIF2 gets released from the preinitiation complex, enabling the large ribosomal subunit to take its place and form a fully assembled active ribosome. Not all AUG codons are equally good at translation initiation. The same codon also serves as a regular methionine codon. In most cases, the very 5' proximal start AUG of the mRNA is recognized as the start. However, numerous exceptions to this rule are shown and named "leaky scanning" [11]. In this instance, the ribosomal preinitiation complex scans through the first AUG site without initiating translation. Instead, one of the downstream AUG codons is selected. Marylin Kozak, a pioneer in translation initiation research, discovered a special consensus sequence surrounding the AUG start codon that defines its probability to be recognized as the start by the ribosome [12, 13]. Consequently, this region was named the "Kozak consensus". It comprises 6-7 nucleotides before AUG and 3 nucleotides after it. The ideal mammalian Kozak sequence look like CGCC(A/G)CC**AUG**GCG, but changes in some nucleotides are possible. More severe deviations from consensus lead to the leaky scanning and affect translational efficiency of the mRNA transcript.

Interestingly, the AUG start codon can be substituted with other triplets, although it dramatically decreases translation efficiency. Most codons with single-base substitution (near-cognate) can initiate translation when positioned in the strong Kozak consensus. Only for AAG and AGG no initiation was detected [14-16]. In case of "near cognate" codons, it is not completely understood which tRNA acts as initiator and which amino

acid it bears. Clearly, some of them can be mispaired with regular initiator methionyl tRNA. Additionally, several studies showed that the leucine tRNA charged with leucine can recognize the CUG codon as the start [17].

It is tempting to speculate that the selection of alternative start codons may serve specific biological purpose. It could produce several variants of a protein from the same mRNA without the need for alternative splicing. Numerous examples have been discovered in yeast and multicellular organisms. For instance, yeast *Ala1* and *Grs1* genes encode the corresponding tRNA synthetases. Non-AUG translation initiation in both of these genes results in longer protein variants bearing a signal for mitochondrial localization [18, 19]. This topic is discussed in greater details in Chapter 3.

1.7 Short upstream reading frames as translational regulators

Most eukaryotic mRNAs encode only a single protein. However, sometimes an additional short open reading frame precedes it. Such ORF is called uORF (upstream) and generally encodes short peptide (less than 10 amino acids). These uORFs are often viewed as potential translation regulators of downstream ORFs. *Gcn4* is one of the best studied examples of such regulation. It is a yeast transcription factor activated by amino acid shortage. There are 4 experimentally verified open reading frames located in its 5' UTR [10, 20]. According to the current model, ribosome scans from the 5' cap of the transcript and, upon encountering the first uORF, initiates translation (Fig. 1.2). Half of the posttermination small ribosomal subunits resume scanning and initiate at the second to fourth uOR, when there are plenty amino acids in the growth medium. Therefore, these uORFs serve as a barrier to ribosomal scanning, significantly depleting the fraction of 40S sububits that would make all the way through from the cap to the *GCN4* start codon.

This regulation gets turned around under conditions of starvation and amino acid depletion, which triggers the phosphorylation of eIF2 by the Gcn2 kinase and, in turn, inhibits the first steps of translation initiation. As a result, the rate of met-tRNA binding to the small ribosomal subunit as it is scanning after termination at uORF-1 drops enough to allow the 40S subunit to bypass all uORFs after the first one and to reach the *GCN4* start codon [10, 20].

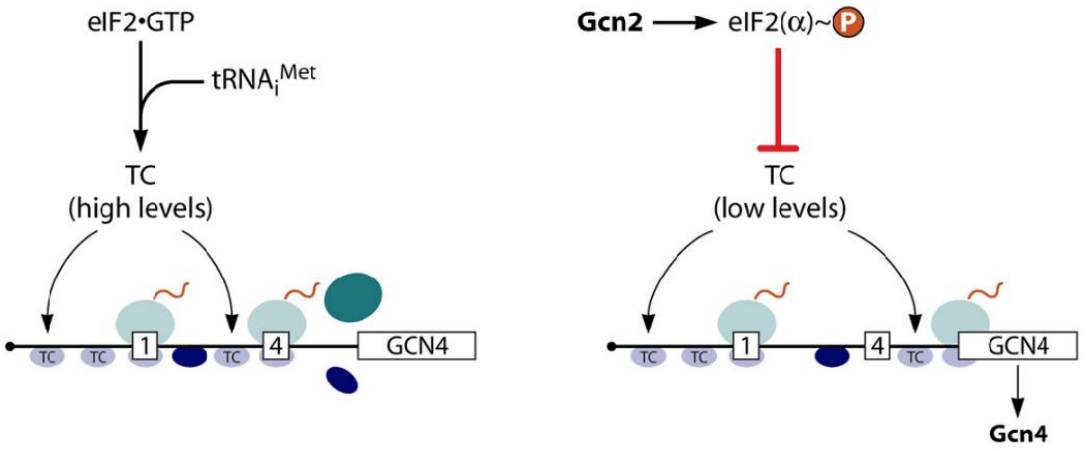

Figure 1.2 Schematic representation of GCN4 regulation by uORFs [10].

Posttranscriptional regulation with uORFs is proposed to have a broader implementation in mammalian cells that express mRNA species having much longer 5'-UTRs. Several well-studied cases are described, such as *ATF4* and *ATF5*. They encode b-ZIP transcription factors like *GCN4* controlling stress-response genes [10, 21].

1.8 Translational regulation of gene expression

Gene expression is broadly regulated at the level of transcription. A single mRNA molecule can be used as a blueprint for multiple copies of a protein. Therefore, adjusting the amount of transcripts allows the cell to control gene expression with fewer resources. Nevertheless, post-transcriptional control is also used frequently. One of the advantages

of regulating gene expression during translation is the reaction speed. Shutting down translation machinery can be achieved much faster than activation of transcription factor that has bind regulatory DNA regions and hence affect transcript synthesis. Global blockade of protein synthesis relies on phosphorylation of a single translation initiation factor. This modification renders it inactive, excluding ribosomal subunits from initiating a new round of translation. Most stresses, such as oxidation, head, and starvation utilize this mechanism at some extent.

Translational control can be limited to a small group of specific transcripts. For instance, some mRNAs bear IRE (iron regulatory element) sites in their sequences [9]. These elements represent a hairpin structure that can be tightly bound by IRP proteins in the absence of iron ions. This complex prevents RNA unwinding by the eIF4A/B complex, so the small ribosomal subunit cannot scan through the IRE hairpin and reach the start codon of a protein. When the concentration of iron increases, it competes with IRE for ERP binding, releasing translational block of related genes [9]. This type of regulation allows maintaining the prompt control of iron concentration in mammalian cells.

Recently, translational control by mTOR (mechanistic target of rapamycin) received a lot of attention. Inhibition of mTOR is believed to extend life span and to be a potent cure for some types of cancer. Mammalian mTOR kinase is a part of two complexes, mTORC1 and mTORC2. Regulation of protein synthesis is a major function of the mTOR1 complex [22]. It regulates the functions of 4E-BP and eIF4G through several intermediate interacting partners, such as S6 kinase. Inhibition of mTOR causes a decrease in global protein synthesis, but unlike the phosphorylation of eIF2, it is achieved

in a radically different way. mTOR was found to regulate translational efficiency of a subset of mRNAs with 5' TOP motif (terminal oligopyrimidine). Many of these mRNAs belong to ribosomal proteins [22]. Therefore, decreasing or increasing their translation efficiency has a direct affect on the number of ribosomes. More specifically, 4E-BPs inhibit translation initiation by preventing the interaction between eIF4E and eIF4G1 initiation factors [22]. As a result, eIF4E no longer binds TOP containing mRNAs suppressing their translation. The described mechanism and TOP motifs are the feature of translation in mammalian cells. Other eukaryotes like yeast, flies, nematodes have no TOP signals, but they have conserved mTOR proteins that also regulate translation.

A special place in translation control is occupied by micro RNAs (miRNAs) – short non-coding RNAs of ~22 nt in length. Although there is some controversy in the extent of regulation achieved by miRNAs, this is currently a rapidly expanding research topic. miRNAs recognize their targets by base pairing. Perfect complementarity between miRNA and mRNA is not required, hence a single miRNA can bind to hundreds of target sites [23]. Many studies in the past few years demonstrated that miRNAs can suppress translation of mRNAs. Mechanistic aspects are obscure, but some miRNAs were shown to interfere with cap recognition by initiation factors as well as to impede joining of ribosomal subunits at the start codon [23]. Most of the identified miRNA binding sites are located at 3' UTRs and may overlap with protein binding sites. Therefore, particular miRNAs compete with RNA binding proteins and alter transcript's cellular localization, in turn changing its translation.

It was reported that, in some rare instances, miRNAs increase translation efficiency of transcripts. They can bind to the 5' UTR regions of transcripts, enhancing translation by an unknown mechanism [23].

This chapter illustrates the complexity of translational control of gene expression. It becomes clear that many aspects of eukaryotic life are regulated at the level of protein synthesis. It is anticipated that this research field would expand rapidly as new high-throughput methods arise.

CHAPTER 2

CUG start codon generates thioredoxin/glutathione reductase isoforms in mouse testes

Note: The results described in this chapter have been published.

Gerashchenko MV, Su D, Gladyshev VN. (2010) CUG start codon generates thioredoxin/glutathione reductase isoforms in mouse testes. *J. Biol. Chem.* 285:4595-4602

2.1. Abstract

Mammalian cytosolic and mitochondrial thioredoxin reductases are essential selenocysteine-containing enzymes that control thioredoxin functions. Thioredoxin/glutathione reductase (TGR) is a third member of this enzyme family. It has an additional glutaredoxin domain and shows highest expression in testes. We found that human and several other mammalian TGR genes lack any AUG codons that could function in translation initiation. Although mouse and rat TGRs have such codons, we detected protein sequences upstream of them by immunoblot assays and direct proteomic analyses. Further gene engineering and expression analyses demonstrated that a CUG codon, located upstream of the sequences previously thought to initiate translation, is the actual start codon in mouse TGR. The use of this codon relies on the Kozak consensus sequence and ribosome scanning mechanism. However, CUG serves as an inefficient start codon that allows downstream initiation, thus generating two isoforms of the

enzyme *in vivo* and *in vitro*. The use of CUG evolved in mammalian TGRs, and, in some of these organisms, GUG is used instead. The newly discovered longer TGR form shows cytosolic localization in cultured cells and is expressed in spermatids in mouse testes. This study shows that CUG codon is used as an inefficient start codon to generate protein isoforms in mouse.

2.2 Introduction

Mammalian thioredoxin reductases (TRs) are essential enzymes that belong to a pyridine nucleotide disulfide oxidoreductase family [24, 25]. In addition to the catalytic site, typical of the entire superfamily, TRs contain a C-terminal penultimate selenocysteine (Sec) residue encoded by UGA codon [26]. This Sec is inserted with the help of Sec insertion sequence (SECIS) element present in the 3'-UTRs of TRs and other selenoprotein genes. The TRs play key roles in the control of cellular redox homeostasis by maintaining thioredoxins (Trxs) in the reduced state, but they are also able to directly reduce certain small molecules such as selenite [27], hydroperoxides [28], dehydroascorbate [29], and NK-lysin [30].

Three TRs exist in mammals: TR1 (also known as TrxR1, TxnRd1 or TrxRα), TR3 (TrxR2, TxnRd2, TrxRβ) and TGR (TR2, TxnRd3). TR1 and TR3 functions are well characterized. The former is a cytosolic enzyme involved in cell growth [31], while the latter is mainly localized to mitochondria and is involved in heart development [32]. Both proteins are present in all vertebrates and are essential for mouse embryogenesis [31, 32]. On the other hand, mammalian TGR is abundant in testis and its function is not well understood [33-37]. It was suggested that TGR promotes disulfide bond isomerisation between GPx4 and other proteins. GPx4 is another selenoprotein abundant in testes,

which is both an enzyme and a structural protein of the mitochondrial sheath in sperm cells [38].

The main feature that distinguishes TGR from other mammalian TRs is an N-terminal glutaredoxin (Grx) domain. Grx is a Trx-fold protein and a component of another major redox system in mammals: the glutathione system [39-41]. Despite an atypical active site motif in the Grx domain (i.e., CxxS instead of CxxC), this domain exhibits Grx activity either in concert with TGR or when expressed alone [34]. Thus, this domain allows TGR to participate in both Trx and glutathione systems [36]. A Grx-containing form of TR1 is also known, but it does not display activities typical of Grx [42, 43]. Prior to TGR discovery, it was thought that Trx and glutathione systems work independently, but increasing evidence suggests crosstalk between these systems. In this regard, several previous observations deserve a particular attention: first, in *Drosophila melanogaster*, the Trx system substitutes for glutathione reductase [44, 45]; second, in *Schistosoma mansoni* and related platyhelminths, there are neither TR nor glutathione reductase (GR), and TGR alone replaces both major redox systems [46-48].

We previously focused on the mouse TGR as a model protein [33-37]. However, examination of its homologs in other mammals revealed a lack of initiation codons in several sequences in the position of the AUG codon previously predicted to serve as the start codon. Translation initiation signals other than AUG are common in viruses; they are also used in bacteria, but are extremely rare in eukaryotes. In mammals, non-AUG triplets with the change in one nucleotide in AUG (with the exception of AGG and AAG codons) could direct translation initiation *in vitro* [49]. However, not all of them are able to serve this function *in vivo*. To date, only about 30 proteins are known that utilize non-

canonical initiation sites in mammals [50]. The majority of these proteins are regulators of transcription and translation, growth factors and cation transport channels. In some cases the utilization of non-AUG codon is driven by IRES structure recognition [51, 52], and in other cases by conventional ribosome scanning mechanism [53, 54]. In this work, we found that CUG is used as a start codon in mouse TGR and that this feature evolved to generate isoforms of this protein.

2.3. Methods

Analysis of TGR genes. Genomic, non-redundant and EST databases at NCBI were scanned with tBLASTN using mouse TGR sequence (NM_153162) as a query. TGR sequences were then extended upstream and aligned using ClustalX.

Expression and purification of recombinant TGR. To generate a construct for expression of the short form of TGR in *E. coli*, cDNA of mouse TGR was amplified using primers F1 and R1 (Table 2.1). The reverse primer contained a SECIS element, derived from *E. coli* formate dehydrogenase H gene, that was inserted immediately downstream of the TAG stop signal of TGR. This PCR product was cloned into pET28a(+) plasmid (Novagen) in-frame with the preexisting N-terminal His-tag using EcoRI and NdeI restriction sites. The construct for expression of the full-length TGR was prepared in two stages. First, the sequence was amplified with primers F2 and R2, then cloned into pET24(+) using EcoRI and NdeI sites. Second, a PCR procedure was used to add a His-tag sequence at the N-terminus using primers F3 and R3. The resulting plasmids were co-transformed into *E. coli* BL21(DE3) cells (New England Biolabs) together with pSUABC plasmid [55]. Cells were grown in LB medium supplemented with 20 μM FAD, 10 μM sodium selenite, kanamycin and chloramphenicol, and

induction of protein synthesis was performed by adding 50 µM IPTG at OD_{600}=1 and incubating cells at 17 ºC overnight.

Table 2.1

Primer	Sequence
F1	5`-CCATATGGCGTCGCCACCCGGCC-3`
R1	5`-TGAATTCTTAGCTAGCGATTGGTGCAGACCTG CAACCGATGGCTAGCCTCAGCAGCCTTTCTG AG-3`
F2	5`-GCATATCGCCATGGAGAAGCCACCGTCCCCG-3`
R2	5`-GGAATTCTTAGGCTAAAGATTGGTGCAGACCTGCAACCGATGTCTAGCCTCA GCAGCCTTTCTGAGTAATGTC-3`
F3	5` ATGGAGAAGCCACCGTCCCCGC-3`
R3	5`-CTGATGATGATGATGATGGCTGCTGCCCATGGTATATCTCCTTCTTAAAGTTA AACAAAATTATTTCTAGAGG-3`
F4	5`CTGGAATTCTGGCCCCGTAGCTGTCCGGTGCGCCCGCGGCCA-3`
R4	5`-ATGGATCCTCAGAAGACGGGCGGCTGG-3`
R5	5`-ATGGATCCGAGAGGCCGCCACGGACAGCAC-3`
F6	5`-GGGATCCACAAAGCCCTGGAGGTGAGCAAGGGCGAGGAGCTGTTC-3`
R6	5`-CGCGGCCGCTTTACTTGTACAGCTCGTC-3`
P1	5`-GACAACAAAGCCCTCGAGAAGCCACC-3`
P2	5`-GGTGGCTTCTCGAGGGCTTTGTTGTC-3`
P3	5`-GACAACAAAGCCATGGAGAAGCCACC-3`
P4	5`-GGTGGCTTCTCCATGGCTTTGTTGTC-3`
P5	5`-GCCCGGAGCGACAACAAAGGTCUGGAGAAGCCACCGTC-3`
P6	5`-GACGGTGGCTTCTCCAGACCTTTGTTGTCGCTCCGGGC-3`
P7	5`-CTCGGCCCGGAGCGACAATTAAGCCCTGGAGAAGCCACCGTC-3`
P8	5`-GACGGTGGCTTCTCCAGGGCTTAATTGTCGCTCCGGGCCGAG-3`
P9	5`-GACGGTGGCTTCTCCAGGGCTTAATTGTCGCTCCGGGCCGAG-3`
P10	5`-GCGGGGACGGTGGCTTATGCAGGGCTTTGTTGTCGCTCCG-3`
P11	5`-CGGCCCGGAGCGACAACAAACATCTGGAGAAGCCACCGTCC-3`
P12	5`-GGACGGTGGCTTCTCCAGATGTTTGTTGTCGCTCCGGGCCG-3`
P13	5`-GTCGCCACCATTGTGAGCAAGGGCGAG-3
P14	5`-CTCGCCCTTGCTCACAATGGTGGCGAC-3`
D1	5`-GGATCCACCGGTCGCCACCAT-3`
D2	5`-AGCCCGGGCGAGGTCTGC-3`
D3	5`-CCGCGTCCCGGTTCCTCAG-3`
D4	5`-GAATTCGAAGCTTGAGCTCGAGATCTGAGTC-3`
D5	5`-GCGGCCGCCTAGGGGAC-3`
D6	5`-GCTCGGCCCGGAGCGAC-3`

Affinity purification of proteins was carried out using Talon resin (Clontech). 50 mM phosphate buffer, pH 7.5, was used containing 300 mM NaCl and 5 mM imidazole as an equilibration/wash solution, and 50 mM phosphate buffer, pH 7.5, containing 300 mM

NaCl and 300 mM imidazole as the elution solution. Following elution, proteins were concentrated for further use.

Constructs for expression in mammalian cells. GFP-fusion constructs were prepared on the basis of pEGFPN1 (Clontech). The N-terminal part of mouse TGR including the extended longer form (designated extTGR) was cloned using primers F4 and R4 into the EcoRI/BamHI sites of pEGFPN1. The sequence located upstream of the previously predicted AUG start codon was separately fused to a GFP sequence using the same forward primer and R5 as a reverse primer resulting in extTGR-GFP construct. Variants of extTGR-GFP plasmid carrying deletions were made as follows: (i) Δ203-256 used primers D1 and D2, (ii) Δ1-92 used primers D3 and D4, and (iii) Δ93-119 used primers D5 and D6 (schematic representation is shown in Fig. 2.1). Plasmids carrying point mutations in mouse TGR Kozak sequence were made as follows: (i) mutation of CUG codon at position 146-148 of cDNA into CTC used primers P1 and P2, (ii) mutation of CUG at position 146-148 into AUG used primers P3 and P4, (iii) mutation of CC at position 144-145 into GT used primers P5 and P6, (iv) mutation of CA in position 136-137 into TT used primers P7 and P8, (v) mutation of GAG at position 149-151 into CAT used primers P9 and P10, and (vi) mutation of GCC at position 143-145 into CAT used primers P11 and P12. To examine TGR expression in HEK 293 cells, we removed the AUG start codon in the extTGR-GFP construct using primers P13 and P14. To replace the AUG codon of GFP in pEGFPN1 with the Kozak sequence of CUG codon in mouse TGR, we used primers F6 and R6 and obtained a PCR product from pEGFPN1, which was then inserted into the same vector digested with BamHI/NotI.

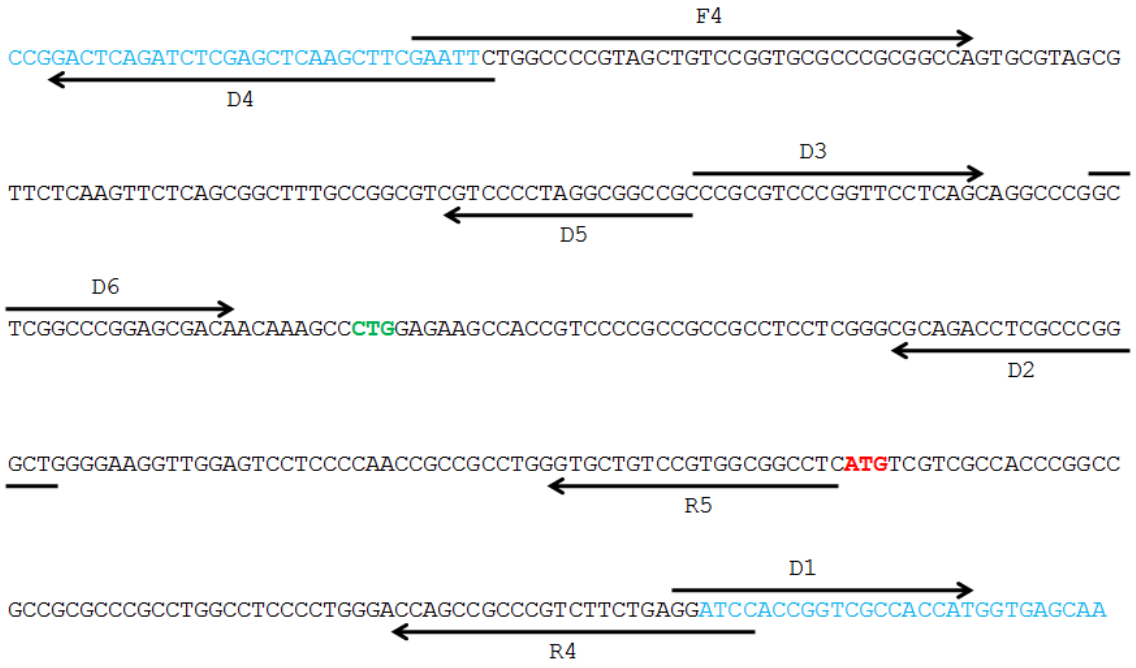

Figure 2.1. Schematic representation of primer positions used for cloning and deletion mutagenesis. Part of the pEGFPN1 vector shown in blue, insertion of mouse TGR sequence (from 2 to 323 nucl.) shown in black. Green and red codons indicate start of TGR-L and TGR, respectively. Black arrows indicate primers, for primer sequences see Table 2.1.

Cell transfection and lysate preparation. Transfection of HEK 293 cells was carried out by calcium chloride method. COS-1 and NIH 3T3 cells were transfected by lipofectamine 2000 (Invitrogen). After 24-48 h of incubation, cells were collected and lysed in CelLytic M (Sigma). Lysates were directly used for SDS-PAGE analysis on 10% Bis-Tris minigels (Invitrogen), followed by western blotting. We prepared rabbit polyclonal antibodies against a shorter version of TGR, and separately against peptide sequences coded by sequences upstream of AUG. Monoclonal anti-GFP antibodies were from Sigma. ECL^{tm} donkey anti-rabbit (or anti-mouse in the case of GFP) IgG horseradish peroxidase-linked antibodies were used as secondary antibodies.

Tissue samples. Testes were taken from C57BL/6 mice fed standard rodent chow (Harlan Teklad, Madison, WI). Tissues were fixed in 10% formalin and processed for paraffin embedding at the Veterinary Diagnostic Center, University of Nebraska-Lincoln. Immunohistochemistry was performed with a Histostain-Plus kit (Zymed) according to the manufacturer's instructions. Briefly, prior to staining, sections of testes were deparaffinized with xylene and passed through a graded series of ethanol. Non-immune goat serum (10%) was used to block non-specific binding. The slides were incubated with antibodies against a short form of TGR (1:300 dilution) or antibodies against N-terminal sequences of TGR (1:10,000 dilution) for 1 h and washed with PBS containing 0.05% Tween 20 (PBST). Biotinylated secondary antibodies were applied to the sections for 10 min. The slides were then washed with PBST and incubated with horseradish-peroxidase-conjugated streptavidin followed by rinsing in PBST. Staining was performed using DAB chromogen. In addition, staining by hematoxylin (Invitrogen, Carlsbad, CA) was done according to the manual. Images were collected using light Olympus AX70 microscope at the University of Nebraska-Lincoln Microscopy Core Facility.

2.4. Results

Several mammalian TGR genes lack AUG start codon. This study began with a surprising observation that human TGR gene lacked an AUG start codon in the position corresponding to the previously predicted start codon in mouse TGR, and that upstream sequences in the human gene lacked any AUG at all. Multiple sequence alignment of mammalian TGR genes revealed that the mouse AUG was only present in rodents and several other animals, such as tupaia and armadillo, whereas humans, other primates and

several other mammals replaced AUG with other codons, and all these sequences lacked AUG upstream in the correct open reading frame (Fig. 2.2).

Mammalian TGRs have coding sequences upstream of AUG codon in mouse TGR. A region upstream of the mouse AUG codon showed high sequence conservation at both nucleotide and protein levels in mammals, and any changes in the nucleotide sequence were multiples of three (i.e., preserving the frame) (Fig. 2.2 and 2.3). This arrangement was indicative of coding sequences.

To directly detect for N-terminal coding sequences, we purified TGR from rat testes and subjected it to LC-MS/MS analyses.

TABLE 2.2

Species name	**Accession number**
Mus musculus	NM_153162 (NCBI nucleotides)
Rattus norvegicus	XM_216204 (NCBI nucleotides)
Tupaia belangeri	AAPY01479941.1 (GenBank)
Dipodomys ordii	1550219844 (NCBI Traces)
Homo sapiens	XM_001130163 (NCBI nucleotides)
Pan troglodytes	XM_516719 (NCBI nucleotides)
Gorilla gorilla	1675688431 (NCBI Traces)
Macaca mulatta	AANU01278373.1 (GenBank)
Papio hamadryas	1964614407 (NCBI Traces)
Callithrix jacchus	1024739016 (NCBI Traces)
Pongo pygmaeus	859560569 (NCBI Traces)
Nomascus leucogenys	1898222911 and 2089571737 (NCBI Traces)
Echinops telfairi	AAIY01742654.1 (GenBank)
Ochotona princeps	1523862870 (NCBI Traces)
Dasypus novemcinctus	AAGV020683171.1 (GenBank)
Bos taurus	1957639188 (NCBI Traces)

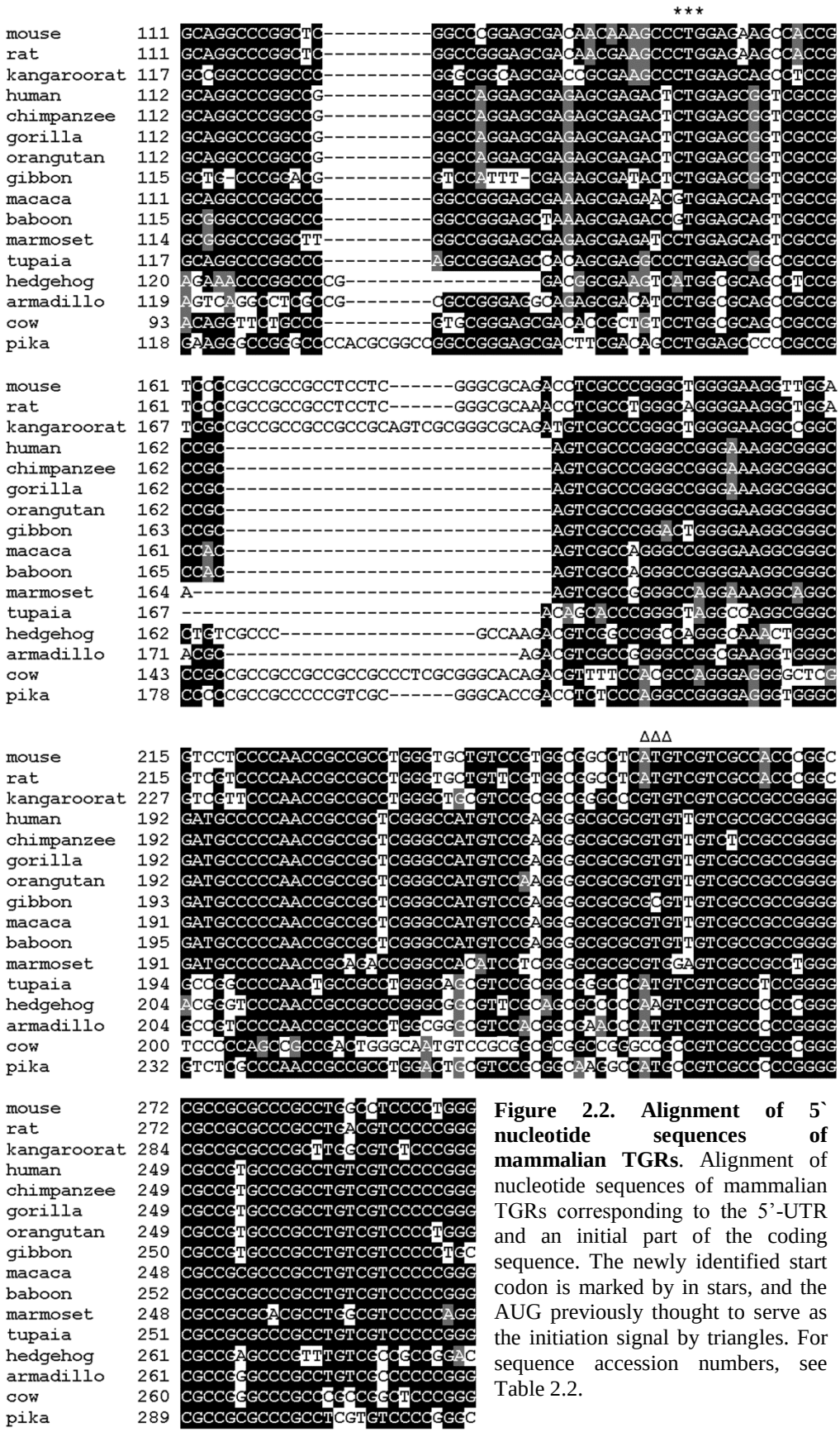

Figure 2.2. Alignment of 5` nucleotide sequences of mammalian TGRs. Alignment of nucleotide sequences of mammalian TGRs corresponding to the 5'-UTR and an initial part of the coding sequence. The newly identified start codon is marked by in stars, and the AUG previously thought to serve as the initiation signal by triangles. For sequence accession numbers, see Table 2.2.

In addition to tryptic peptides corresponding to internal sequences of TGR, this procedure identified a peptide that extended 8 amino acids upstream of the AUG codon. Taken together, these observations suggested that non-canonical translation initiation is used in mammalian TGR genes.

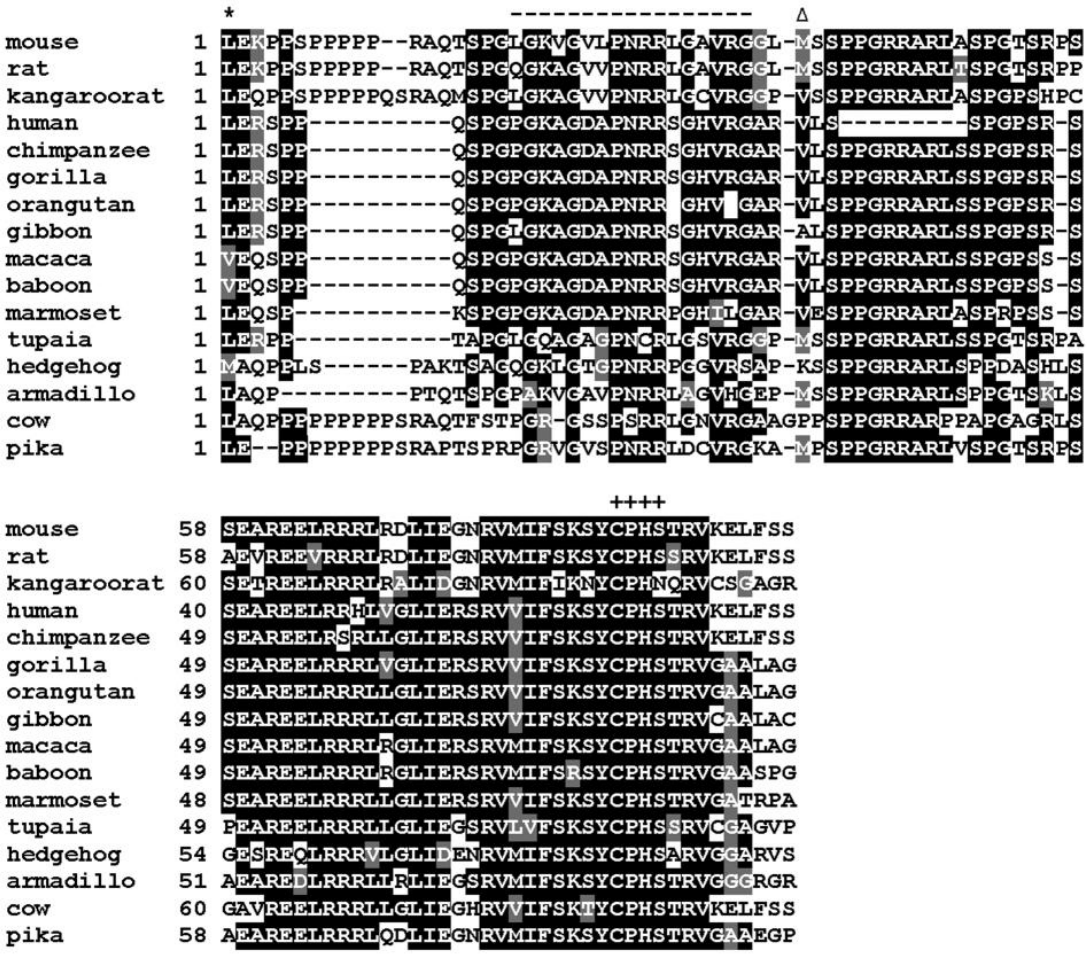

Figure 2.3. Alignment of N-terminal protein sequences of mammalian TGRs. Alignment of N-terminal sequences of mammalian TGRs, starting from the CUG codon. First amino acid residue in protein is marked by star, and the methionine previously thought to be the initial residue by triangle. The active site of the Grx domain of TGR is designated by crosses. The peptide in the longer TGR form that was used as antigen for polyclonal antibodies is marked with dashes.

Mouse TGR has an alternative start codon upstream of the previously reported AUG. We verified the presence of the actual TGR mRNA sequences in testes of C57BL/6 mice

by amplification and cDNA sequencing. Then, we cloned a region of mouse TGR corresponding to 22 amino acids downstream of the AUG together with the entire upstream sequences, or only the sequences upstream of the AUG, and prepared fusion constructs with GFP (including its AUG start codon). These constructs were then transfected into HEK 293, NIH 3T3 or COS-1 cells and examined for translation initiation by subjecting protein extracts to western blots with anti-GFP antibodies (Fig. 2.4 A,B). When the sequences located upstream of the AUG were used, two protein forms were detected. One was GFP alone and the other was a fusion of GFP with sequences coded by the upstream region; these sequences must have corresponded to the N-terminus of TGR. The other construct yielded three bands which corresponded to GFP alone, AUG-originated mouse TGR and the form that begun with the natural start codon of GFP. Simultaneous detection of multiple protein forms suggested that the natural upstream start codon is inefficiently used.

Mammalian TGRs have a candidate CUG start codon. To identify an upstream start codon, we introduced deletions in the mouse TGR cDNA sequence and transfected such constructs into HEK 293 cells. When nucleotides 203-256 were deleted, the upper band, which corresponded to translation from an alternative start codon became lower (Fig. 2.5 A). This observation suggested that the upstream start codon should be closer to the 5` end. Removing nucleotides 1-92 of TGR cDNA did not affect translation initiation. Nucleotides 93-119 that are conserved in mammals were not required either (Fig. 3B). From these observations, and taking into account a 24 nucleotide gap in rat and mouse sequences relative to human TGR, we defined a 40 nucleotide region that contained an alternative start codon (Fig. 2.2). Based on the sequence alignment, CUG codon at

position 146-148 emerged as a promising candidate for translation initiation. This codon was preserved in almost all mammals except hedgehog and gibbon, but in those organisms, it was replaced with GUG or AUG codons, two most common translation start sites in nature.

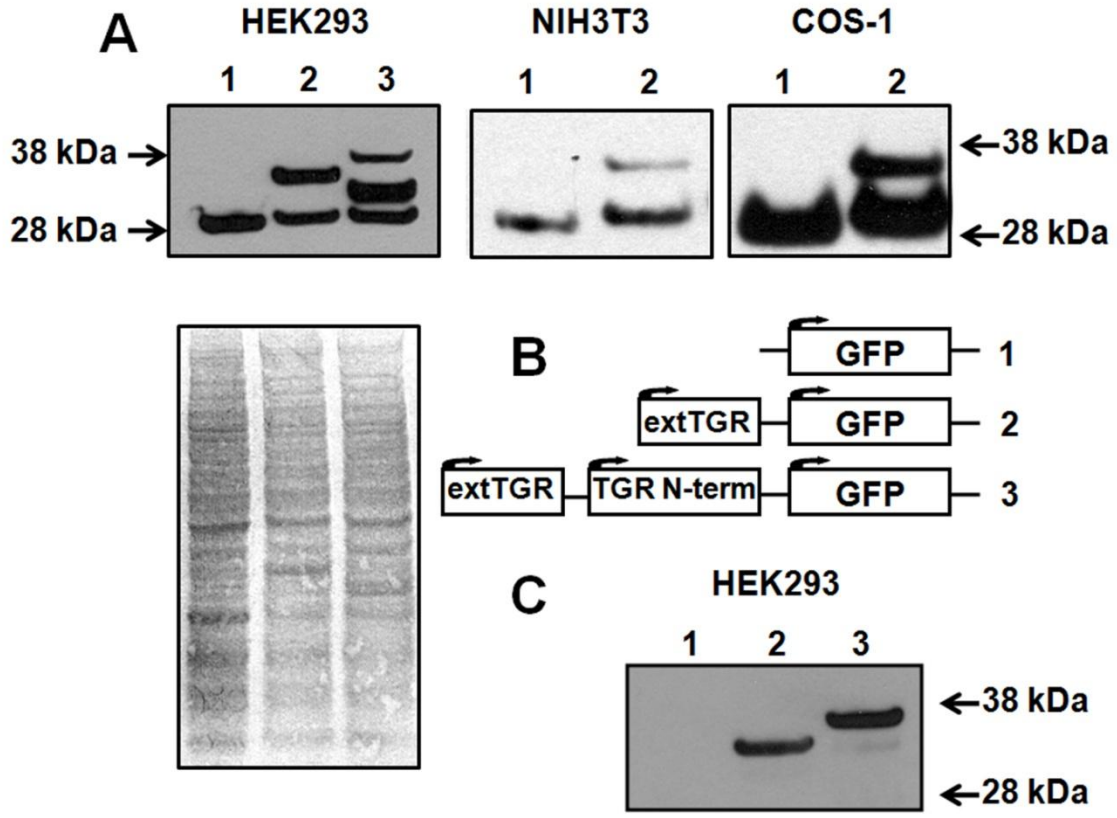

Figure 2.4. Expression of TGR constructs in cell lines. A, immunoblot analysis of expression products. The samples were as follows: lane 1, GFP alone; lane 2, GFP coding sequence fused with the 5`-UTR and sequences coding for the N-terminal region of TGR; lane 3, GFP sequences fused with the 5`-UTR and sequences coding for an additional N-terminal 22 amino acids. The constructs were expressed in indicated cells lines and the expressed proteins were probed in immunoblot assays with antibodies specific for GFP. B, schematic representation of constructs used in A. Numbers correspond to those in A. Black arrows indicate positions at which translation initiation occurs. C, imminoblot analysis of samples shown in A with antibodies specific for the N-terminal region of TGR-L.

To test if CUG is the natural start codon, we mutated it into CUC. The mutation resulted in the loss of the larger protein form, whereas the forms that started from

downstream sequences were intact (Fig. 2.5 A, lane 5). When CUG was changed to AUG, efficiency of translation initiation from this site increased such that only the larger protein form was detected (Fig. 2.5 A, lane 6). Thus, CUG is a natural, albeit inefficient, start codon; and in addition, a downstream AUG codon can serve as a start codon wherein two protein forms are synthesized from the same mRNA. The novel large TGR form is hereafter designated as TGR-L; it differs from the previously known TGR form by a 4 kDa N-terminal extension.

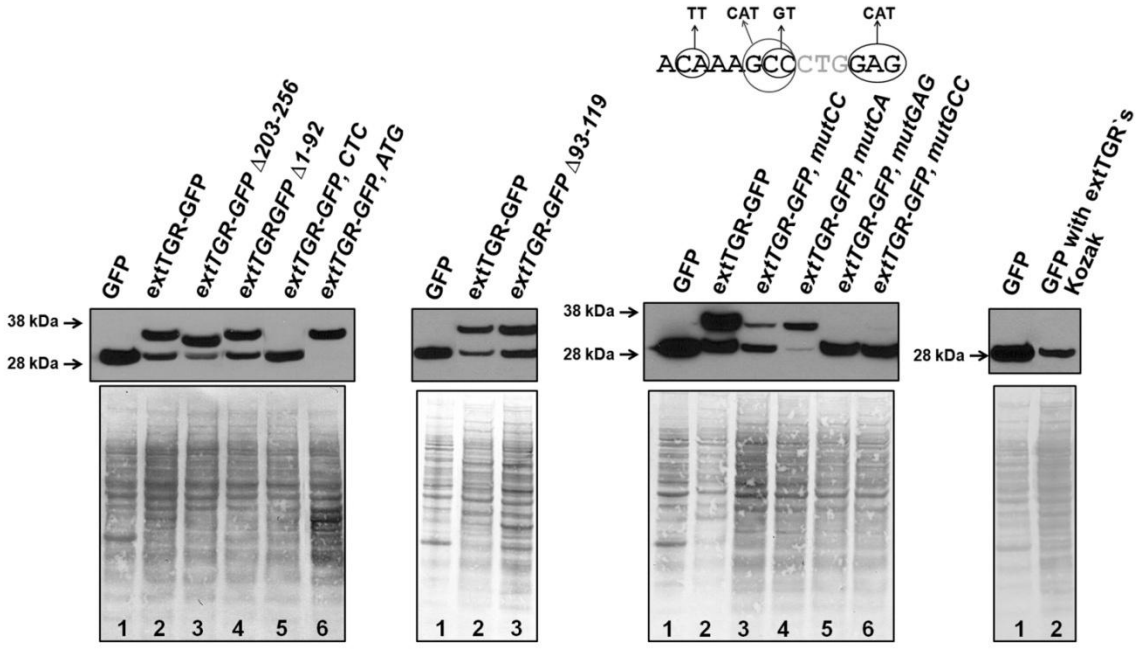

Figure 2.5. Analysis of translation initiation function of the CUG codon. A, deletion and point mutation of CUG codon and its flanking areas were used to examine the ability of CUG to initiate translation. Lane 1, GFP control; lane 2, GFP fused with 5`-UTR and sequences coding for the N-terminal region of TGR-L; lane 3, GFP fusion construct carrying deletion at nucleotide positions 203-256; lane 4, GFP fusion construct with deletion at nucleotide positions 1-92; lane 5, GFP fusion construct in which CUG was mutated to CUC in nucleotide position 148; lane 6, GFP fusion construct in which CUG codon was mutated to AUG. B, deletion of conserved sequences upstream of CUG does not affect translation initiation. Lane 1, GFP control: lane 2, GFP fusion construct that contains the 5`-UTR and sequences coding for the N-terminal region of TGR-L; lane 3, same construct, but with deletion in nucleotide positions 93-119. C, Kozak consensus sequences flanking the CUG codon are important for efficient initiation of translation. Upper panel summarizes mutations that were examined. Middle panel shows an immunoblot analysis with anti-GFP antibodies. Lysates of HEK 293 cells were used.

Cells were transfected with GFP fusion constructs with mutations in the Kozak consensus sequence of the CUG codon. Lane 1, GFP control; lane 2, GFP fusion construct coding for the N-terminal part of TGR-L; lane 3, the same construct but CC was replaced with GT; lane 4, CA was replaced with TT; lane 5, CAG was replaced with CAT; and lane 6, GCC was replaced with CAT. Lower panel shows protein loading control. D, Kozak consensus sequence of the longer form of TGR is sufficient to initiate translation at CUG codon in HEK 293 cells. Immunoblot analysis with antibodies specific for GFP is shown. Lane 1, expression of control GFP construct in HEK 293 cells; lane 2, expression of the GFP construct with AUG start codon replaced by CUG and its Kozak sequence (-8 to +6 positions relative to CUG) derived from the longer form of TGR.

Taking into account the data shown in Fig. 2.4 and 2.5, we developed polyclonal antibodies against LGKVGVLPNRRLGAVRG peptide, which is part of the N-terminal extension and is unique to the long TGR form. Western blot analysis of the same samples as those shown in Fig. 2.4A with these antibodies provided additional evidence for TGR-L existence (Fig. 2.4C). The upper band in Fig. 2.4A corresponded to the translation of the N-terminal part of TGR-L. Despite excellent Kozak sequence of the CUG codon, it serves as a weak initiator of translation and is subject to leaky scanning. This explains the observation of a middle band in lane 3 that originates from AUG start codon of the short TGR form. Moreover, this AUG has a weak Kozak consensus; thus, ribosome may also initiate translation at the downstream GFP sequence.

Mechanism of translation initiation from the CUG codon. All cases of non-canonical start codon usage in mammals can be separated into two groups: IRES-dependent and IRES-independent. We carried out site-directed mutagenesis studies to determine the mechanism of CUG-initiated translation: specifically, whether it utilizes IRES or based on a typical ribosome scanning mechanism. As discussed above, deletion of sequences upstream or downstream of the CUG codon and its 25 flanking nucleotides at the 5` end had no influence on translation initiation. Since the shortest experimentally verified viral

IRES has a length of 56 nucleotides, while the average size in mammals is about 300 nucleotides according to the IRESite database [56], the functional sequences flanking CUG could not accommodate IRES. We further examined this mRNA region by Mfold and did not identify a stable mRNA structure that could function as IRES. Thus, IRES-dependent mechanism is not likely. We also made a set of constructs with point mutations in the consensus sequence that flanks the CUG (Fig. 4C). This region is referred to as the Kozak sequence for alternative initiation in a recent bioinformatics study (27); positions -7, -6, and -4 are particularly important in addition to the classical Kozak. These mutations either completely blocked or severely inhibited translation initiation. On the other hand, certain changes in nucleotide sequences facilitated it (e.g., replacement of GA in positions -7 and -6 with TT increased fidelity of the CUG start codon). To further exclude a role of possible vicinal sequences in the CUG-driven translation initiation, we replaced the native AUG start codon in the control GFP construct by the Kozak region (from -8 to +6 positions) of TGR-L, including the CUG codon, and expressed this construct in HEK 293 cells. A clear and sharp band was observed by western blotting with anti-GFP antibodies. Thus, the Kozak sequence of mouse TGR is sufficient for translation initiation at CUG codon (Fig. 2.5 D).

Tissue distribution of TGR and TGR-L. Previous studies showed that TGR is abundant in testis and is expressed in seminiferous tubuli. To examine TGR-L expression *in vivo*, we employed polyclonal antibodies against the unique N-terminal part of TGR-L. As a control, we used antibodies prepared against a recombinant TGR which lacked the N-terminal 4 kDa region; these antibodies recognized all TGR forms (total TGR). Immunohistochemical analyses revealed that the total TGR was evenly distributed among

seminiferous tubuli cells, while TGR-L was less abundant on the outer edge of tubuli (Fig. 2.6). Thus, the long TGR form was expressed in mouse testes and it showed an expression pattern that differed from that of total TGR. Overall, both TGR forms were apparently present in mouse testes.

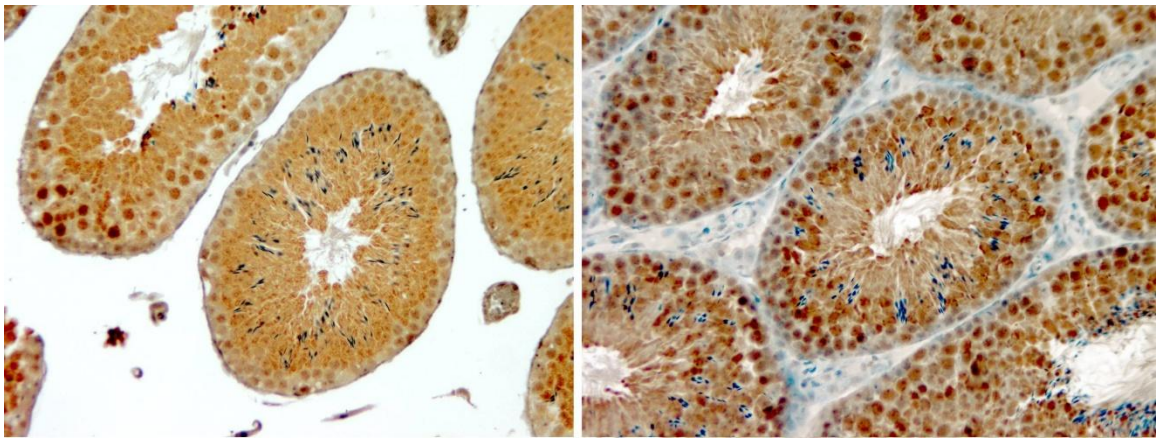

Figure 2.6. **Localization of TGR in mouse seminiferous tubuli.** The left panel shows staining of a mouse testis section with antibodies specific for the N-terminal part of TGR-L (brown) and hematoxylin (blue). The right panel shows staining of a mouse testis section with antibodies specific for the recombinant TGR and hematoxylin.

Localization of TGR in cultured cells. We examined the N-terminal sequence of TGR-L for being a localization signal. Computational analyses by PSORT II and other programs did not identify signal sequences in this region of TGR. We transfected HEK 293 cells with the construct coding for the 4 kDa sequence of TGR-L (designated as extTGR in the figures) followed by GFP. Since the CUG start codon was not stringent in translation initiation resulting in background from unfused GFP, we mutated the natural AUG start codon of GFP. The resulting protein product was expressed and localized to cytosol (Fig. 2.7). As a control, we used a similar vector with the N-terminal part of regular TGR. Some nuclear staining was also detected, but it was not different from the expression of GFP alone (which has an inherent ability to pass through the nuclear

membrane). The data suggest that the N-terminal sequence of TGR-L is not involved in targeting the protein to cellular compartments when expressed in transfected cells.

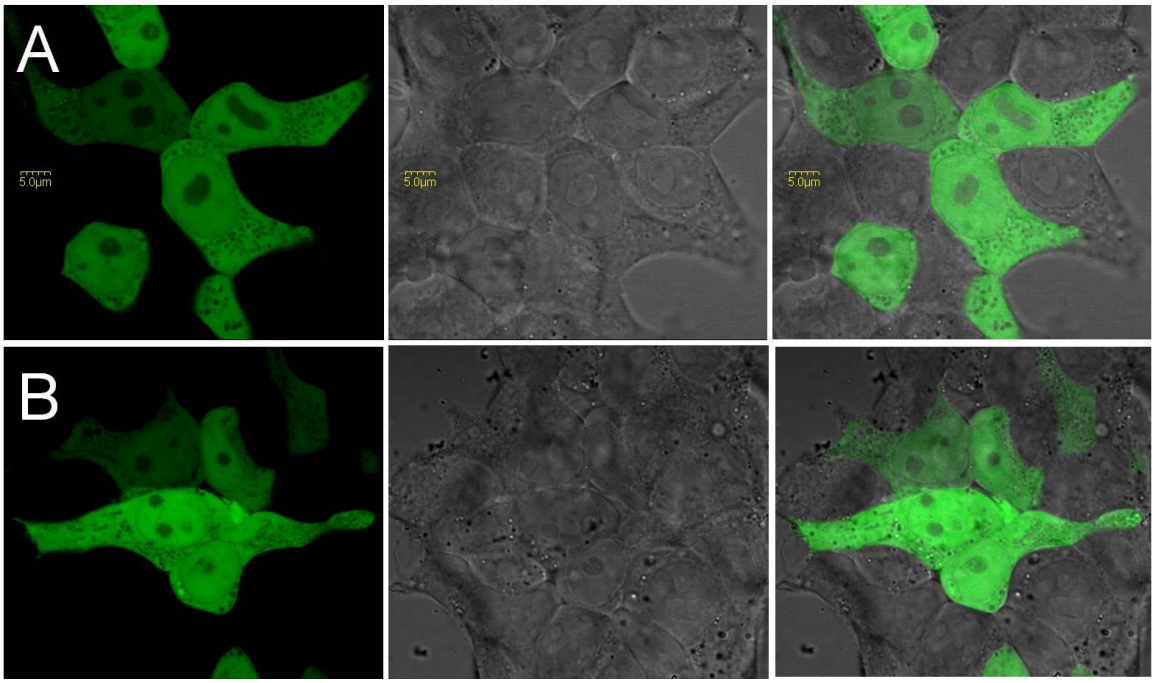

Figure 2.7. Localization of the N-terminal region of TGR-L ant a shorter form of TGR fused to GFP in HEK 293 cells. A, Expression construct coding for the N-terminal part of TGR-L (up to the nucleotide position 256) fused to GFP. B, Expression construct coding for the N-terminal part of short TGR (nucleotides 256-323) fused to GFP.

The two forms of TGR occur in mammals. We examined if the two forms of TGR occur *in vivo*. Western blot analysis of mouse testes with antibodies specific for recombinant TGR revealed two bands, which corresponded to TGR-L and a shorter form (Fig. 2.8). The assignment of the upper band to TGR-L was further verified by western blotting with antibodies specific for the synthetic peptide in the N-terminal portion of TGR-L. We also observed heterogeneity of TGR forms in rat testes (data not shown). Overall, the two TGR forms were both generated upon expression of the gene in cell culture experiments and existed *in vivo*.

CUG usage evolved in mammals. We traced the use of the CUG codon in TGRs by analyzing evolution of this protein in vertebrates and examining translation initiation signals in these sequences. Placing this information on the tree of life revealed that the ancestral form of TGR had a AUG start codon. This codon is still used in fish, birds, and amphibians. Most mammals, however, contain extension at 5` end of TGR mRNA with a conserved CUG codon, and some primates (baboon, macaca) use GUG instead. The acquired use of CUG and its conservation in mammals indicate that translation initiation is best served by a non-canonical codon in TGRs.

2.5. Discussion

Numerous isoforms of mammalian thioredoxin reductases TR1 and TR3 are known. Some of them were predicted from the analyses of EST sequences, and some were experimentally verified [43, 57-61]. As a rule, alternative first exon splicing is used in these genes wherein different mRNA forms are transcribed from unique promoters and generate unique N-terminal sequences which converge onto the common TR module. TR1 and TR3 forms may be targeted by their N-terminal sequences to different cellular compartments or to distinct interacting partners. In some cases, an alternative form of TR1, normally a cytosolic protein, localizes to mitochondria, whereas the mitochondrial TR3 can be targeted to the cytosol when alternative first exon splicing skips over the sequences coding for the mitochondrial signal [57]. There is also an intriguing isoform of TR1 that has a Grx domain that shows no activity in the Grx assays. However, this domain can be activated by mutating two amino acids in the active site. Further studies linked the function of the Grx-containing TR1 to cytoskeleton rearrangements and cell shape [62, 63].

For TGR, however, no alternative forms were described. Searches in the EST database did not reveal obvious candidates. But on western blots, mouse and rat TGRs appear as two bands (Fig. 2.8).

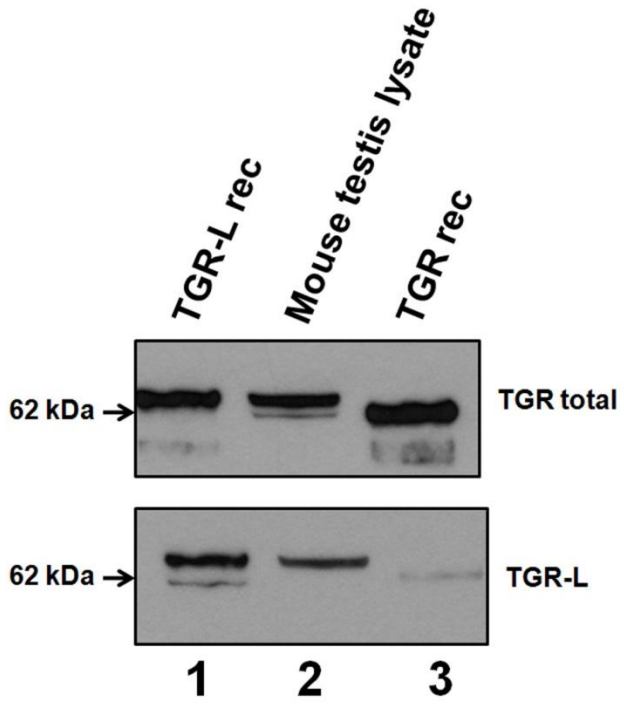

Figure 2.8. TGR isoforms in mouse testes. Mouse testis lysate and recombinant TGR forms were subjected to immunoblot analyses. Upper panel represents samples analyzed with anti-TGR antibodies (lane 2). Recombinant TGR (lane 3) and TGR-L (lane 1) were purified from *E. coli*. Lower panel shows the same membrane probed with antibodies specific for the N-terminal part of TGR-L.

Through a series of experiments, we found that TGR exists in two forms generated by leakage through a weak translation initiator codon. In experiments involving HEK 293 cells, translation from a AUG codon was more efficient than from a CUG codon. But in mouse testis this may not be the case. In addition, the mechanisms and details of translation initiation may vary between various organisms. In any case, the mechanism used to generate TGR forms differs from that in other TRs, although all three enzymes are present during spermatid development (cytosolic and mitochondrial TRs are essential enzymes that are expressed in all cell types). Both TGR isoforms occur in mouse testes; however, the longer form appears to be more abundant. Analysis of the alignment of

mammalian TGRs suggests that human and some other organisms may only have a single form, the long form of TGR, due to absence of the AUG codon downstream the CUG in mouse and rat sequences.

Utilization of non-AUG start codons is highly unusual in mammals. Only several proteins are known to use codons other than AUG, and they include growth factors, kinases and transcription factors [50]. Some of these proteins evolved an IRES structure which facilitates translation initiation, while others still utilize cap-dependent, ribosome-scanning mechanism. For instance, FGF2 (fibroblast growth factor) has as many as five vicinal CUG codons, four of which participate in the IRES-driven translation while the function of the remaining one is cap-dependent [64]. In rare cases, regulation of CUG translation by trans-acting factors can occur [65]. Recent studies suggest an unknown CUG-specific methionine-independent translational mechanism [66]. In the majority of such proteins, a non-AUG codon is auxiliary to the main AUG start signal and is located upstream of it. Several such proteins are transcriptional regulators known to use only non-AUG codons to initiate translation [67-69]. One exception is a phosphoribosylpyrophosphate synthase which is involved in nucleotide synthesis and was found exclusively in testes [70].

The selective usage of CUG codon in TGR generates protein isoforms, and evolutionary analyses suggest that this function evolved specifically in mammals and has been almost uniformly preserved in these organisms. Non-mammalian (e.g., amphibians, fish, birds) TGRs still use AUG as the start codon. However, in the majority of mammals, a conserved CUG is used instead. Some primates replace it with GUG, which perhaps could also serve as an inefficient start signal. Our data suggest that the function of CUG

is to provide inefficient translation initiation that allows production of two forms of TGR from a single mRNA species.

CHAPTER 3

Genome-wide ribosomal profiling reveals complex translational regulation in response to oxidative stress

Note: The results described in this chapter have been published.

Gerashchenko MV, Lobanov AV, Gladyshev VN. (2012) Genome-wide ribosomal profiling reveals complex translational regulation in response to oxidative stress. *Proc. Nalt. Acad. Sci USA.* 109(43):17394-9

3.1 Abstract

Information on unique and coordinated regulation of transcription and translation in response to stress is central to the understanding of cellular homeostasis. Here we used ribosome profiling coupled with next-generation sequencing to examine the interplay between transcription and translation under conditions of hydrogen peroxide treatment in *Saccharomyces cerevisiae*. Hydrogen peroxide treatment led to a massive and rapid increase in ribosome occupancy of short upstream ORFs, including those with non-AUG translational starts, and of the N-terminal regions of ORFs that preceded the transcriptional response. In addition, this treatment induced the synthesis of N-terminally extended proteins and elevated stop codon read-through and frameshift events. It also increased ribosome occupancy at the beginning of ORFs and potentially the duration of the elongation step. We identified proteins whose synthesis was regulated rapidly by hydrogen peroxide posttranscriptionally; however, for the majority of genes increased protein synthesis followed transcriptional regulation. These data define the landscape of genome-wide regulation of translation in response to hydrogen peroxide and suggest that

potentiation (coregulation of the transcript level and translation) is a feature of oxidative stress.

3.2 Introduction

Gene expression may be controlled at multiple levels. Globally, it is regulated by histones and satellite proteins. Locally, promoters, enhancers and other regulatory elements are used to guide transcription. Numerous studies yielded datasets involving the networks of transcription factors and described the associated mechanisms of transcriptional regulation. Developments in microarray technology facilitated such studies and made them affordable for individual laboratories. Accordingly, a vast number of studies emerged that describe transcriptional responses to various treatments, stimuli, knockouts, etc. Conversely, the investigation of the regulation of gene expression at the level of translation lagged behind due to lack of accessible high-throughput methods.

It is often assumed that changes in mRNA abundance are proportional to changes in protein synthesis in the cell, but numerous exceptions are known. One powerful approach to directly assess changes in protein abundance is the use of whole proteome mass spectrometry, but this method is inferior to mRNA profiling in its throughput and can only detect a fraction of protein products in the cell [71]. Other high-throughput approaches are also available, such as fluorescent protein reporter libraries [72-74]. However, they are designed for quantification of individual proteins rather than for addressing the details of translation. Indirect approaches, such as microarray profiling of mRNAs within monosomes and polysomes, are popular as well [75-78]. These methods enable estimation of the mRNAs transcripts that are being translated. Recent advances in next generation sequencing enhanced data acquisition, improved sensitivity and made

this method superior to microarrays in its throughput [79]. Importantly, it allowed examining mRNA abundance and protein translation in the same sample with high accuracy (with subcodon resolution) [80, 81]. This experimental strategy involves deep sequencing of mRNA fragments (footprints) buried inside the actively translating ribosomes. Protein translation can be inferred from footprint abundance. Coupled with regular mRNA sequencing (mRNA-seq) analyses, these data give information on actual mRNA sequences that are being translated, identity of the reading frames used, and ribosomal density at each position within these mRNAs. Hereafter, we refer to this method as ribosome profiling or Ribo-seq. Another promising application of Ribo-seq is measuring of translational regulation by monitoring translation efficiency (TE), which is the amount of footprint normalized to underlying mRNA abundance.

In the current study, we applied Ribo-seq to investigate the fine details of *Saccharomyces cerevisiae* response to oxidative stress, caused by hydrogen peroxide treatment. A key advantage of this method is a much higher sensitivity compared to microarrays. With this method, we were able to detect changes in transcription and its regulation within 5 min of treatment. Oxidative stress is one of the best studied regulators of transcription [82], but little is known about how this stress changes protein abundance and post-transcriptional regulation. Previous studies pointed to a weak correlation between transcriptional and translational gene responses, i.e., elevated mRNA transcripts in stressed cells did not match the set of proteins that changed abundance. Microarray analyses revealed that only 15% of genes involved in translational response showed the corresponding changes at the mRNA levels [76]. Our study focused on using Ribo-seq to examine precisely translation and its regulation by oxidative stress.

3.3 Methods

Preparation of Lysates. Hydrogen peroxide (0.2 mM final concentration) was added to 400 mL of yeast culture, and the culture was incubated further for either 5 or 30 min. A 50 mL aliquot was taken rapidly and pelleted by centrifugation for 1 min at 3,400 × g at 4 °C; then the pellet was frozen immediately in liquid nitrogen. This aliquot was used for mRNA isolation. The rest of the yeast culture was treated with 0.1 g/L cycloheximide, incubated for 3 min with shaking, and centrifuged at 3,400 × g for 4 min. The pellet was resuspended in 3 mL of ice-cold polysome lysis buffer [20 mM Tris·HCl (pH 8.0), 140 mM KCl, 5 mM $MgCl_2$, 0.2g/L cycloheximide, 1% Triton-×100] and recentrifuged. The supernatant was removed, and the pellet was treated with 1.2 mL of the polysome lysis buffer along with an equal amount of glass beads. The resulting mix was vortexed rigorously five times for 1 min with 1-min breaks. The aqueous fraction was collected and clarified by centrifugation for 10 min at 20, 000 × g. The final yeast lysate containing intact ribosomes was flash frozen in liquid nitrogen.

Ribosome Fractionation and RNA Extraction. A 50 U aliquot of the cell extract (OD_{260}) was treated with 1,000 U of Escherichia coli RNase I (Ambion) and incubated for 1 h at room temperature with gentle shaking. The sample volume was brought to 1 mL by adding polysome gradient buffer [20 mM Tris·HCl (pH 8.0), 140 mM KCl, 5 mM MgCl2, 0.2g/L cycloheximide, 0.5 mM DTT]. Sucrose gradients (10–50% wt/wt) were prepared in SW41 ultracentrifuge tubes (Beckman) using a freeze-thaw method [83]. RNase digested and control samples were loaded onto gradients and spun for 3 h at 35,000 rpm and 4 °C in a SW41 rotor (Beckman). Gradients were fractionated at 1 mL/min using the Brandel gradient fractionation system coupled with the BioRad UV

detector, which continually monitored OD_{254} values. As a chase solution, 60% (wt/wt) sucrose was used, and fractions representing the monosome peak were pooled in one tube. Each sample was filtered through an Amicon-100 microcentrifugator (Millipore) for 10 min at 10,000 × g. The release buffer [20 mM Tris·HCl (pH7.0), 2 mM EDTA, 40 U/mL Superase-In (Ambion)] was added to the retentate until the volume reached 0.5 mL, and each sample was incubated further for 10 min on ice and then was filtered again. Flow-through fractions containing the majority of footprints were collected, and RNA was purified by hot acid phenol extraction and precipitated by ethanol with glycogen as a coprecipitant. Pellets were solubilized in 10 μL of water and analyzed on 15% Tris/borate/EDTA TBE-Urea polyacrylamide gels (Invitrogen). The bands around 28–32 nt were cut off, and RNA was eluted in 300 μL of the elution buffer containing 20 mM Tris·HCl (pH 7.0), 2 mM EDTA, 0.5 M ammonium acetate, and 2 μL Superase-In, precipitated, and resuspended in 8 μL of water. After addition of 1 μL of T4 kinase "A" buffer and 1 μL of T4 kinase (Fermentas), the mixture was incubated for 60 min at 37 °C, inactivated for 5 min at 80 °C, and ethanol precipitated.

Library Construction for Footprint Sequencing. Polyadenylation of RNA footprints was performed by adding 0.5 U of poly(A) polymerase (New England Biolabs) in a total volume of 5 μL and incubating the mixture for 15 min at 37 °C. The enzyme was inactivated by heating the mixture at 80 °C for 10 min. The whole reaction mix was used for reverse transcription. Superscript III (Invitrogen) polymerase was used according to manufacturer's instructions in a total reaction volume 12 μL. The RT-library primer was used for each individual sample. Finally, 0.5 μL of 2M sodium hydroxide was added to hydrolyze RNA from RNA-DNA duplexes, and the sample was incubated for 30 min at

98 °C. Then, 0.5 μL of 2 M HCl was applied to neutralize the solution. Upon the addition of an equal volume of TBE-sample buffer (Invitrogen), the reverse-transcription mixture was loaded onto a 10% TBE-urea gel (Invitrogen). The band corresponding to the elongated RT-library primer was cut, and DNA was eluted in 300 μL of 20 mM Tris·HCl (pH 7.0). An important step for efficient enrichment of ribosomal footprints was the subtractive hybridization of contaminating rRNA fragments. For this step, the biotinylated DNA oligonucleotide "bioAntiRiboPrime" (Table 3.1) was attached to streptavidin-activated magnetic beads (Invitrogen) as recommended in the manufacturer's manual. Ribosomal footprints eluted from the gel were incubated with these beads, and nonribosomal fragments that did not bind to the beads were collected and ethanol-precipitated. They served as substrates for CircLigase II (Epicentre) in a 10-μL reaction mix. Circularized ribosomal footprints were used as a template for the final library amplification step. PCR conditions were set as follows: 0.5 μL of Phusion polymerase (New England Biolabs), 1 μL of 10 mM dNTP, 1 ul of DNA template, 10 μL of HF buffer (New England Biolabs), and 10 pmol of custom ill-Cluster3 and ill-Cluster4 primers compatible with Illumina sequencers (Table 3.1) in a 50-μL mixture. Annealing took place at 70 °C for 15 s, and elongation took place at 72 °C for 10 s. Several reaction tubes were set up to be removed from the PCR machine after 12–18 cycles. The product yield was analyzed on 8% nondenaturing TBE polyacrylamide gels to select samples (based on PCR conditions) before the appearance of nonspecific bands. The library was cut from the gel, eluted in 20 mM Tris·HCl (pH 7.0), ethanol-precipitated, and sequenced on the Illumina GLx2 or HiSeq2000 platforms.

TABLE 3.1. List of primers used in library preparation.

RT-library*	pGATCGTCGGACTGTAGAACTCTθCAAGCAGAAGACGGCATACGATTT TTTTTTTTTTTTTTTTVN
RT library1**	pCGTGATGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCC GTATCATT/iSp18/ CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTVN
RT library2**	pTGGTCAGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCC GTATCATT/iSp18/ CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTVN
RT library3**	pATTGGCGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCC GTATCATT/iSp18/ CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTVN
RT library4**	pCTGATCGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCG TATCATT/iSp18/ CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTVN
Ill-cluster 3	CAAGCAGAAGACGGCATACGA
Ill-cluster 4	AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAG
Ill-cluster 5	AATGATACGGCGACCACCGA
bioAntiRiboPrime	bio\GAGGTGCACAATCGACCG

* - P - phosphate; θ - abasic site (dSpacer); iSp18 – internal spacer 18.
** - primers with a barcode tag at the 5`-end.

mRNA Extraction. Frozen aliquots were thawed and lysed in 400 µL of lysis buffer (mRNA DIRECT kit; Invitrogen). A 250-µL aliquot of magnetic beads and two rounds of purification were implemented according to the manufacturer's protocol.

mRNA Sequencing Library Construction. mRNA was fragmented by alkaline solution [2 mM EDTA, 100 mM Na_2CO_3 (pH 9.2)], the fragments were loaded onto a 15% TBE-urea gel, and the 28- to 32-nt region was cut from the gel. Further steps in library preparation were identical to those used for ribosomal footprints, the only difference being that barcoded RT-library 1–4 primers were used that allowed multiplexing of samples for sequencing (Table 3.1). The subtractive hybridization step was omitted. The PCR annealing temperature was set to 60 °C with ill-Cluster3 and ill-Cluster5 primers.

Bioinformatics Analyses. In-house Perl scripts were used to prepare reference databases. We created several references using the Saccharomyces Genome Database as a starting point. The largest reference ("Functional") included all cDNAs except for

transposons and dubious genes. Among these cDNAs, the genes with a high degree of sequence similarity were combined into single records. This dataset was used for differential gene-expression and translation studies. Additionally, 100 nt from the 5′ end of each gene were deleted to avoid bias caused by the region with elevated footprint density. Another reference ("noRepeat") included only unique gene sequences to which footprints could be aligned unambiguously. It was used when the nucleotide position-sensitive features of translation were examined. Alignment of sequencing reads was performed by Bowtie software v.0.12.7 [84] allowing two mismatches per read. Alignment against 5′ UTR was done with one mismatch allowed. Because every read bears a poly(A) tail at the end, we omitted all "A" from the 3′ ends of sequences before aligning. Reads shorter than 23 nt after polyA removal were discarded.

Calculation of Translation Efficiency. Translational efficiency (TE) is a measure of how well translated a particular gene is relative to its mRNA abundance. TE can be defined as the number of footprints divided by the number of mRNA-seq reads normalized to gene length and total number of reads, i.e., footprint in reads per kilobase per million mapped reads (rpkm)/mRNA rpkm. A higher TE value represents greater potency of mRNA for translation. TE was used to examine translationally regulated genes. If a gene had a log2 (TE change) above 1.5 or below 1.5, it was considered up- or down-regulated, respectively. Fig. 3.1 shows the fraction of false positives at the selected threshold.

Inferring Translation Rate from Sequencing Data. Sequenced footprints represent pieces of mRNA trapped in the active translating ribosomes. A higher number of footprints aligned to a gene sequence implies a higher yield of the corresponding protein.

This assumption is more reliable for genes with more even footprint coverage. Significant deviation from evenness may indicate ribosomal pauses in certain locations; such pauses complicate the inference of protein production. In this study, we observed higher density of footprints at the beginning of mRNAs; therefore, we discarded 100 nucleotides from the 5′ end of every gene to minimize unevenness of footprint coverage along transcripts.

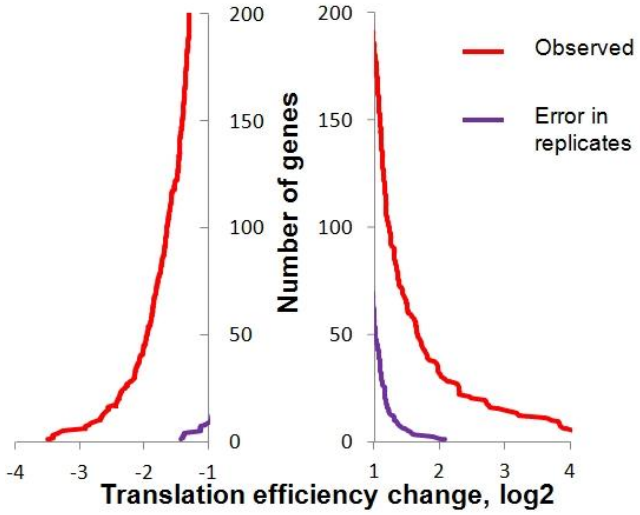

Figure 3.1. **Estimation of the error rate for TE change**. The purple line shows how many genes would be mistakenly assigned as affected at a certain threshold if two biological replicates are compared. The red line shows a number of genes with the TE changed between an initial state and following 30 min peroxide treatment.

Differential Gene Translation Analysis. All experimental samples were collected in duplicate. Based on the correlation between the replicates, we set up an rpkm threshold of 10 for the genes whose translation and transcription could be determined reproducibly (Fig. 3.2 A and B). The gene was considered regulated if its rpkm value changed more than 2.6-fold (1.4 in log2 scale). This threshold eliminated most of false-positive hits (Fig. 3.3).

Comparing Translation Changes with Transcription Changes. In an ideal situation, assuming that transcript abundance is the only determinant for protein translation, changes in transcript abundance would be followed by the same changes in footprint abundance. In reality such coordinated changes never happen, as illustrated in Fig. 3.12

B. Axis values are calculated as footprint change versus transcript change between two experimental conditions. Footprint change is defined as log2[(Footprints in peroxide-treated sample, rpkm)/(Footprints in initial sample, rpkm)]. Transcript change is defined in a same way for mRNA-seq reads.

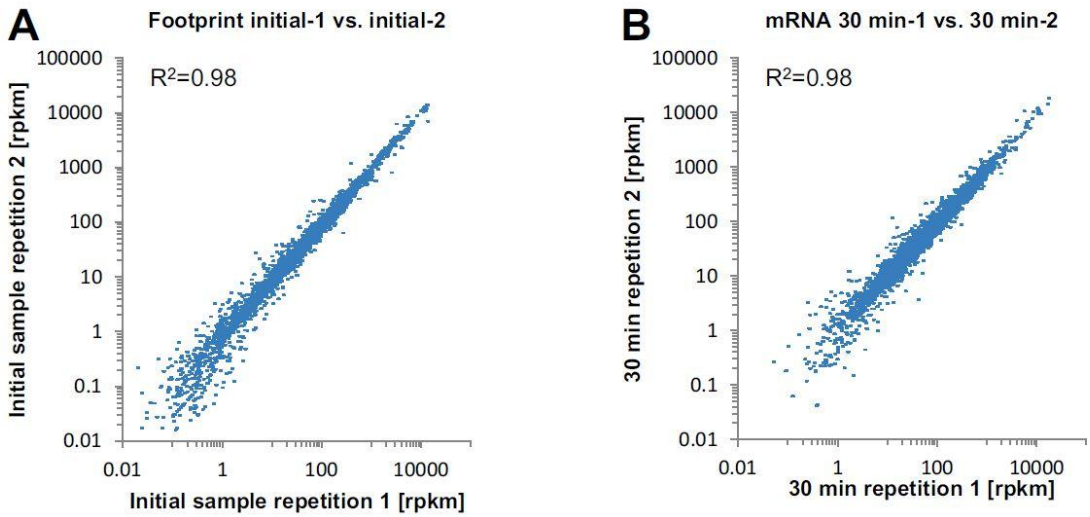

Figure 3.2. Comparison of gene expression between two replicates of footprints. Panel A shows footprints, and panel B mRNA reads. Correlation coefficients are indicated in the figure

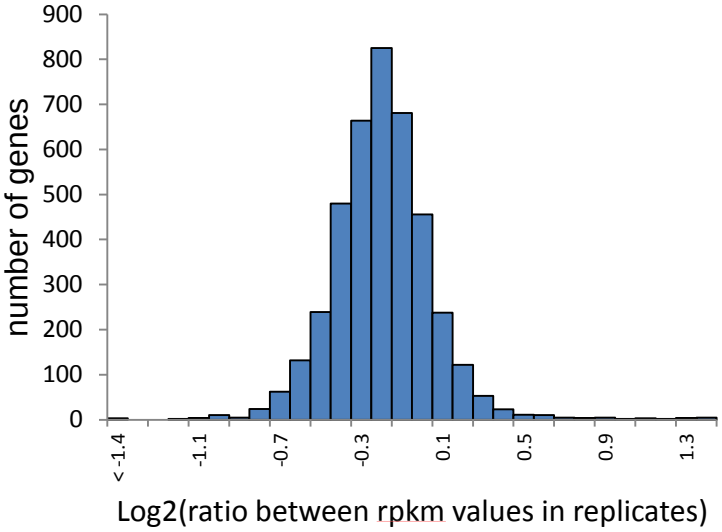

Figure 3.3. Justification for threshold selection. The majority of differences between the two replicates fit in +/- 1 interval on the log2 scale. However, to minimize false-positive hits, we set up the +/- 1.4 interval as the threshold. It allowed us to avoid most

false positives in the 5 min peroxide treatment samples, in which the overall count of regulated genes was low.

Codon Translation Analysis. In an ideal situation, ribosomal footprints should be 28 nt in length. However, RNase I, which was used to degrade unprotected mRNA segments, occasionally left extra nucleotides or cut off extra nucleotides. By plotting a distribution of the footprint length, we found that RNase creates footprints mostly are 27–29 nt in length (Fig. 3.4).

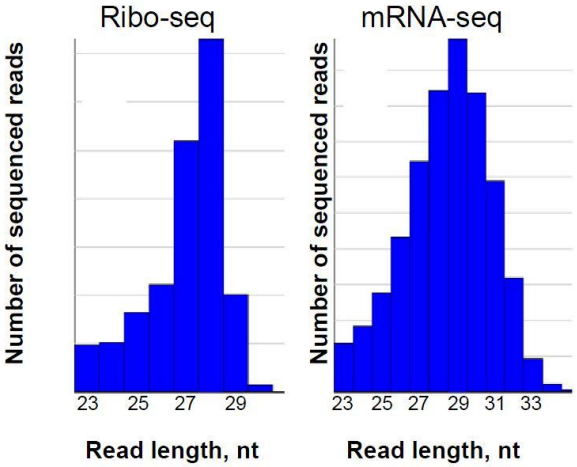

Figure 3.4. Distribution of sequence reads by length in the control sample. Left panel shows footprints, right panel mRNA reads. Poly(A) tails were omitted from the reads.

A footprint can be aligned to the reference ORFs, and the position of its 5′ end relative to the reading frame can be obtained. If the 5′ end of a footprint matched the exact border of a codon, we considered it "ideal." If the 5′ end of a footprint matched the position of a codon ±1 nt, we deleted or added the first nucleotide, respectively. Thus, we minimized the error of ribosome position determination and defined which codon was located in the A site. To estimate differences in TE among various codons (61 codons in total), we used following procedure. First, predicted occupancy was calculated for each type of codon as its frequency in mRNA sequence, normalized to gene expression (translation) and length (assuming that all codons are translated at the same rate). These

values were compared with the observed frequencies. As a measure of difference, we used the following formula [(Observed) − (Predicted)]/(Predicted), which gave us an estimate of how the use of a particular codon compared with the predicted value.

Frameshift Analyses. The regions 50 nt downstream of stop codons of every gene were examined for the presence of ribosomal footprints. Footprint mapping similar to gene-coverage analysis was used to select possible frameshift extensions over readthrough events. Footprint reads were assigned to all possible reading frames and counted. During counting, reads were used as is; i.e., we did not add or subtract nucleotides from the 5′ ends. Candidates with signs of translation in different frames downstream of their stop codons were checked manually to exclude dubious cases and to define the frameshift regions more precisely.

Selecting Proteins with Potential N-Terminal Extensions. Some genes have ribosome profiling (Ribo-seq) footprints mapped to their 5′ UTRs in close proximity to annotated start codons. We marked proteins as potential bearers of N-terminal extensions if they satisfied three conditions. First, they were represented by at least 50 rpkm Ribo-seq counts 45 nt upstream of known ORFs. Second, the majority of Ribo-seq footprints mapped to these regions were in the same reading frame as the annotated proteins. Third, there were no stop codons in this frame 45 nt upstream of the annotated start codon.

3.4 Results

An overview of the Ribo-seq method that we used to examine the regulation of translation by oxidative stress is given in Fig. 3.5 A. Each translating ribosome protects ~28 nucleotides on the translated mRNA, and the unprotected regions are removed by subjecting mRNAs to RNase I digestion. The protected mRNA pieces (footprints) are

extracted and analyzed by deep sequencing. Because their length is known, the exact codons that occupy the A and P sites of the ribosome can be determined. This information is used to identify frameshifts, read-through events, and altered codon use. Additionally, quantification of footprints provides an opportunity to estimate changes in translation for every mRNA species. A key factor that decreases throughput of this method is that only 5% of total yeast RNA consists of mRNA in rapidly growing yeast cells [85]. Previously, contamination was eliminated during footprint preparation by ultrafiltration, which is not very efficient; i.e., the fraction of ribosomal RNA fragments in sequencing libraries approached 80%, with an average value of about 60%, as observed in previous studies [80] and our own pilot experiments. To improve the throughput of the method, we examined the content of contaminating rRNA fragments. In our footprint samples a particular fragment of the 28S ribosomal subunit was responsible for 90% of contamination. An additional step of subtractive hybridization allowed us to get rid of this specific fragment, and 95% of the resulting library consisted of mRNA footprints (Tables 3.2 and 3.3). Such high purity made possible sample multiplexing, which increased throughput and decreased cost.

Oxidative Stress Increases Ribosome Occupancy of Upstream ORFs. Upstream ORFs (uORFs), short ORFs immediately upstream of the main gene sequence, are known to modulate gene expression in response to amino acid depletion and other types of stress. One of the best-studied examples is the regulation of GCN4, which has multiple uORFs that block its translation when sufficient levels of amino acids are present but allow translation when amino acids are depleted [20]. Precise mapping and thorough characterization of such uORFs have been complicated because of the lack of sensitive

methods. Bioinformatics analysis and modeling were used instead [86]. Ribo-seq overcomes this challenge, detecting uORFs quantitatively and mapping them to the mRNA at a single nucleotide resolution [80].

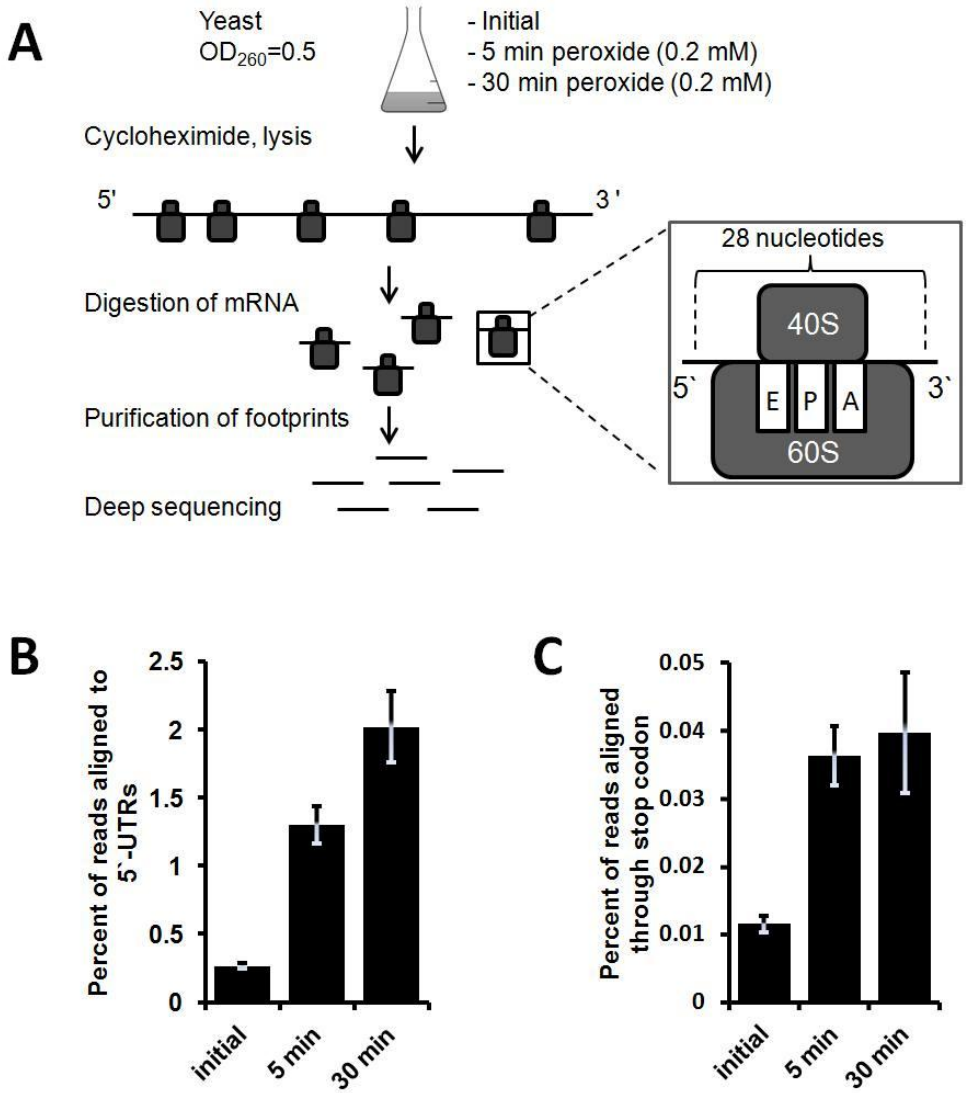

Figure 3.5. **Oxidative stress affects fidelity of translational machinery**. (**A**) Design of the experiment. See text for details. (**B**) Hydrogen peroxide treatment leads to an increase in 5′-UTR translation. Yeast cultures were treated with 0.2 mM hydrogen peroxide for 5 or 30 min. Untreated cells served as control. A fivefold increase in net translation of 5′ UTRs occurred after 5 min of incubation. Incubation with hydrogen peroxide for 30 min further increased 5′-UTR translation. (**C**) Oxidative stress leads to translation read-through events at stop codons. Experimental conditions are as in B. Error bars indicate SEM. Measurements from biological replicates are shown.

TABLE 3.2. Statistics of deep sequencing reads

Footprints	*Initial-1*	*initial-2*	*5min-1*	*5min-2*	*30min-1*	*30min-2*
total reads	27,145,924	84,852,974	13,341,052	82,763,853	5,981,943	80,589,116
Genomic, non-rRNA	25,302,082	79,522,848	12,204,639	74,177,834	5,271,843	70,444,698
ORF_minus100nt, uniq	18,690,126	61,222,201	8,297,207	49,006,214	3,435,799	42,568,826
5' UTR	61,769	228.496	176,003	867,375	120,516	1,241,515
mRNA	*Initial-1*	*5min-1*	*5min-2*	*30min-1*	*30min-2*	
total reads	22,560,757	18,283,784	13,424,316	20,910,828	19,871,495	
Genomic, non-rRNA	20,707,193	17,434,262	12,398,186	18,250,816	19,301,893	
ORF_minus100nt, uniq	12,211,073	9,849,232	7,614,102	11,834,969	12,257,517	
5' UTR	297,592	361,129	257,098	298,010	319,222	

TABLE 3.3. Proteins with translated N-terminal extensions

Gene	*Name*
YDR077W	SED1
YHR179W	OYE2
YHR087W	RTC3
YPL154C	PEP4
YDL022W	GPD1
YIR037W	HYR1
YBR221C	PDB1
YGL039W	
YDR086C	SSS1
YKL103C	LAP4
YBR121C	GRS1
YOR039W	RPS12
YMR297W	PRC1
YNL064C	YDJ1
YJL183W	MNN11
YFR049W	
YGR146C	ECL1
YMR088C	VBA1
YPL183W-A	TAE4
YKL004W	AUR1
YKR052C	MRS4
YIL124W	AYR1
YDR043C	NRG1
YAL012W	CYS3
YKL138C	MRPL31
YER048W-A	ISD11
YER133W	GLC7
YPL170W	DAP1
YJL099W	CHS6
YOR335C	ALA1
YPR182W	SMX3
YLR332W	MID2

We first used Ribo-seq to examine if oxidative stress caused by hydrogen peroxide treatment affects the diversity and abundance of uORFs. We used annotated 5′ UTRs from the yeast transcriptome- sequencing study [87]. Among them, surprisingly many UTRs (1,800 genes) showed detectable presence of translating ribosomes at the uORFs. These uORFs often overlapped with each other and frequently lacked AUG start codons. In many cases, this observation complicated the analysis of individual uORFs; i.e., often it was unclear if a single uORF or several adjacent uORFs were present in the gene. uORFs are thought to be short, but when clustered they may occupy long sequences upstream of actual ORFs. Thus, we call such regions "upstream translation islets." They can be short or long, represent a single uORF or an uORF cluster, and change their length and composition in response to various treatments. To quantify the translation events within 5′ UTRs, we assigned sequencing reads to the entire 5′ UTRs rather than attempting to separate them into individual uORFs.

We next compared yeast cells treated with 0.2 mM hydrogen peroxide for 5 or 30 min with corresponding untreated cells. Even short (5-min) incubation resulted in a fivefold increase in the ribosomal footprints aligning to the 5′ UTRs (Fig. 3.5B). We detected 847 5′ UTRs whose coverage by footprints increased more than 2.6-fold under these conditions, and the 30-min treatment increased this number to 1,217 UTRs. Interestingly, the changes in 5′ UTR utilization generally were more pronounced than those of downstream genes and occurred at an earlier time point. In addition, the majority of uORFs initiated translation at non-AUG codons under both normal conditions and oxidative stress, as is seen in cells under conditions of amino acid depletion [80].

Interestingly, translation of 5′ UTRs increased uniformly during stress, and no 5′ UTR was down-regulated under these conditions.

Many Genes Show Translation Immediately Upstream of Their Known Start Codons. Analyzing uORF distribution, we observed multiple translation events immediately upstream (i.e., within 45 nt) of their AUG start codons, and oxidative stress increased these events significantly. Elevated ribosome occupancy at uORFs may be caused by slower elongation or, conversely, by increased translation. Up-regulated translation can lead to one of two possible outcomes. First, the translation upstream of AUG may correspond to the N-terminal extensions of some proteins. Second, uORFs in the vicinity of start codons could influence the translation of downstream genes. They may facilitate reinitiation of the ribosome at a downstream AUG codon because the distance between the uORF's stop codon and the following start codon is short (10–15 nt on average). On the other hand, dissociation of the ribosome complex at the uORF stop codon could prevent translation of the main gene. Supporting the first possibility, our analysis revealed five strong candidates with N-terminal extensions in untreated samples, 13 in samples treated with peroxide for 5 min, and 32 in samples treated for 30 min (Table 3.3). These peptides were translated in the same reading frame as the downstream gene and usually started with a non-AUG codon. Fig. 3.6 features proteins selected to represent different scenarios of the N-terminal extension/ORF interplay. The only two known yeast proteins with N-terminal non-AUG extensions, ALA1 and GRS1 tRNA synthetases [18, 19], were among our identified proteins. In these two proteins, N-terminal sequences serve as signal peptides, directing a fraction of these proteins to mitochondria. We examined the subcellular localization of our detected protein

candidates using Gene Ontology (GO) annotation of the SGD database. Twenty-one of 32 proteins had experimentally verified localization in both cytosol and another compartment, such as mitochondria, Golgi, vacuoles, and membranes. Such an enrichment of GO terms supports the idea of regulation by targeted protein localization in response to oxidative stress.

At the genome-wide level, the majority of 5′ UTRs supported uORF translation rather than N-terminal protein extensions. We observed intricate and widespread translation of 5′ UTRs under conditions of oxidative stress. Some common cases are shown in Fig. 3.7, illustrated by four representative proteins. Remarkably, the coverage profiles for every gene were alike in different experimental conditions and were nearly identical in replicates.

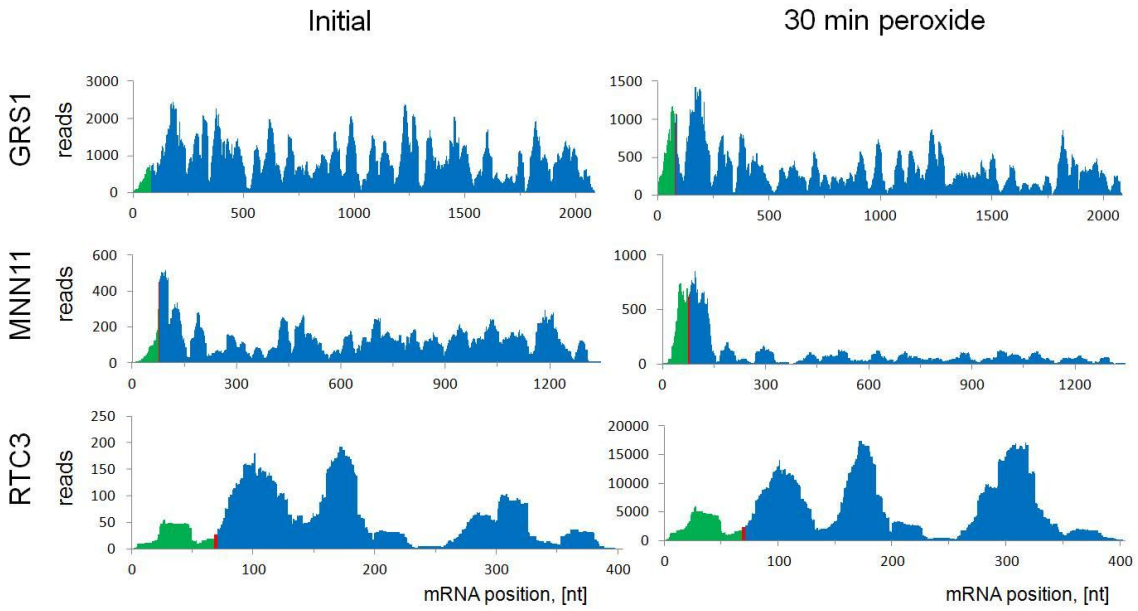

Figure 3.6. **Examples of proteins with N-terminal extensions**. These proteins were selected from Table S3 to represent different scenarios of the N-terminal extension/ORF interplay. This figure compares the translation of proteins in control and 30-min hydrogen peroxide treatment samples. The entire 5′ UTR and ORF of the gene were used to generate the coverage density map. The 5′ UTR part of the mRNA is shown in green, the AUG start codon in red, and the annotated gene in blue.

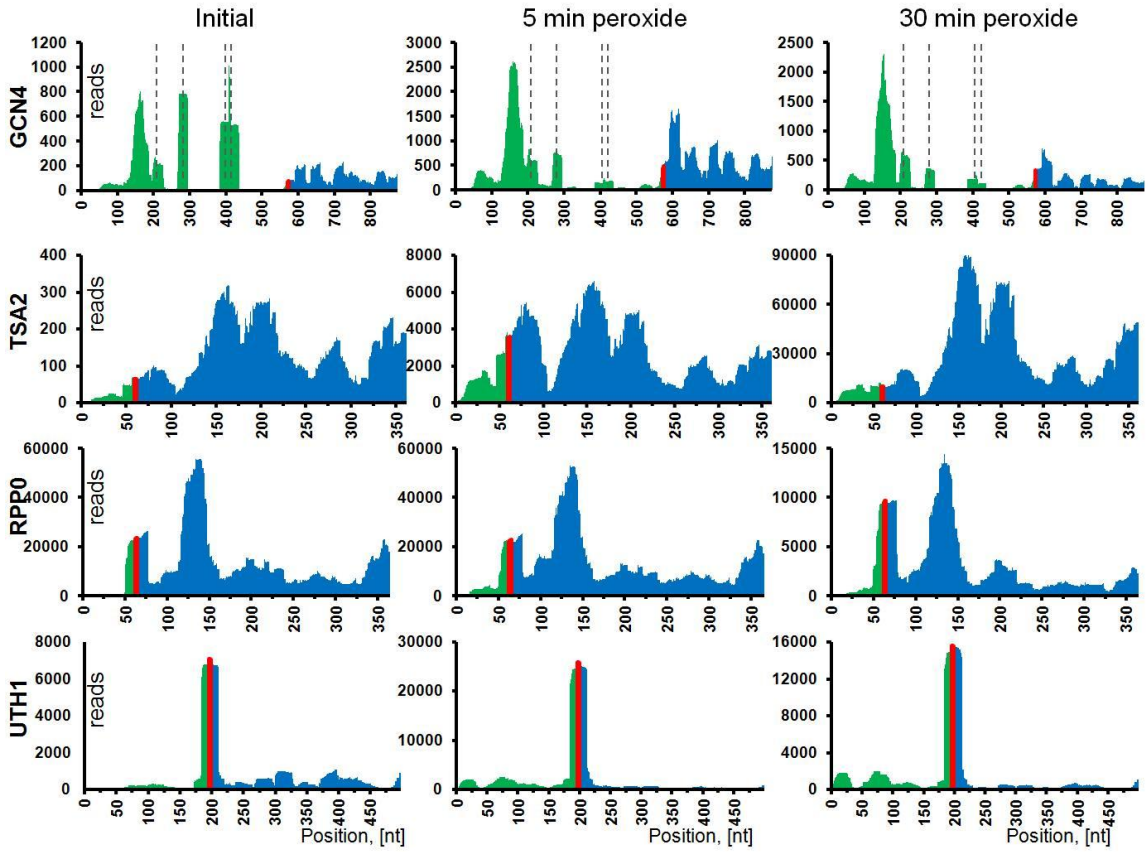

Figure 3.7. Examples of 5′-UTR translation during oxidative stress. Ribosome footprint coverage for four different mRNAs discussed in the text illustrates various patterns of translation. Panels show the footprint coverage of certain mRNAs with no inframe stop codons upstream of annotated genes. For each mRNA, translation following 5- and 30- min hydrogen peroxide treatment is given. Untreated yeast cells served as a control. The entire 5′ UTR and 300 nt of the gene sequence were used to generate the coverage density map. The 5′-UTR part of the mRNA is shown in green, the AUG start codon in red, and the annotated gene in blue. Dashed lines in GCN4 graphs indicate positions of known uORFs.

Oxidative Stress Induces Translational Read-Through of Stop Codons and Frameshifting. Oxidative damage is known to impact ribosomal proteins and translation factors. We examined the rate of readthrough events at stop codons. Termination of translation appeared to be very efficient in the control sample, based on poor read coverage of 3′ UTRs immediately downstream of stop codons (Fig. 3.5 C). Oxidative stress increased read-through events threefold in both 5- and 30-min samples. We also developed a simple method for frameshift search and validation that is technically similar

to the search for N-terminal extensions. A short region downstream of the stop codon for each annotated gene was examined for the presence of ribosomal footprints with coverage comparable to the gene itself. A handful of candidates were confirmed manually. For validation, the 5′ ends of footprints aligned to the regions upstream or downstream of the known frameshift were quantified and assigned to the matching reading frame. The frame with the highest count would correspond to the actual ORF. This approach is shown in Fig. 3.8 and Fig. 3.9 for two known frameshifts in *S. cerevisiae*, antizyme and protein ABP140, respectively [88]. Further analysis of genes for read-through of annotated stop codons yielded four additional genes with+1 frameshifts (i.e., ribosome slipping one nucleotide towards 3' end) (Table 3.4). An example is shown in Fig. 3.8B. All these frameshifts were detected under conditions of oxidative stress.

Table 3.4. Proteins with frameshifts induced by hydrogen peroxide treatment

Gene	*Name*
YKL157W	APE2
YPL224C	MMT2
YJR103W	URA8
YLR179C	Function unknown

Correlation Between Transcriptional and Translational Responses to Oxidative Stress. In *S. cerevisiae*, ~1,700 genes are regulated by hydrogen peroxide at the level of transcription, including ~900 genes of the environmental stress response cluster, which encompasses genes regulated in response to various stresses such as heat shock, starvation, and oxidative stress [82]. Next-generation sequencing technologies can improve the sensitivity and dynamic range of gene expression analysis significantly. We found that after 5-min treatment with hydrogen peroxide transcriptional changes were observed for 116 genes, of which 10 were down-regulated and 106 were up-regulated.

The 30-min treatment yielded transcriptional changes in 1,497 genes (529 down-regulated and 968 up-regulated) with the threshold of 2.6-fold (see Datasets S1, S2 which could be downloaded as online supplement to this manuscript, and Fig. 3.10 for comparison of mRNA-seq with microarrays from ref. [82]).

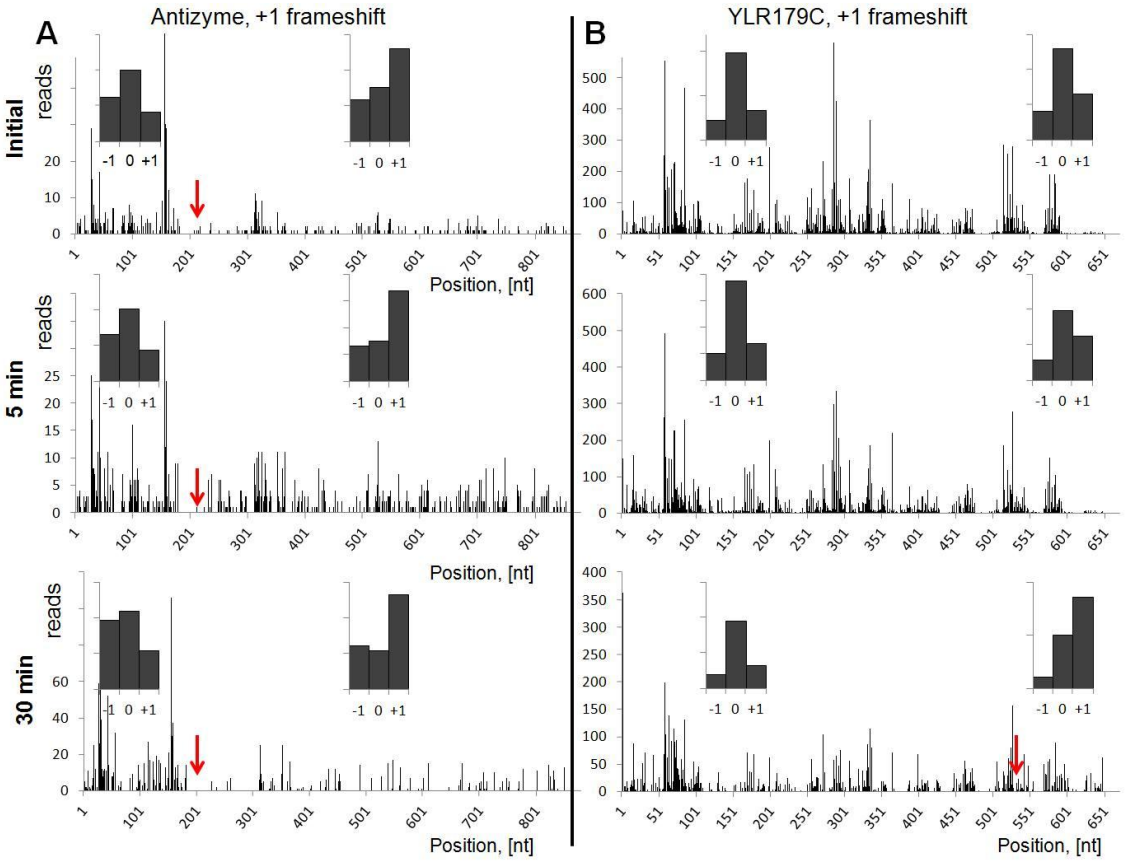

Figure 3.8. **Ribo-seq allows identification of frameshifts (red arrows).** (**A**) Validation of the known frameshift in the antizyme gene. (**B**) Oxidative stress leads to a frameshift in the product of the YRL179C gene. We observed a change of frame, leading to translation of a longer protein in the 30- min peroxide treatment sample. The 5′ ends of footprints were mapped to the genomic sequence of YRL179C. (Insets) Histograms show the count of footprints, matching one of three possible frames either to the left or to the right of the frameshift. The "0" frame is the one with the annotated start codon. The highest count of footprints matched the "0" frame before the frameshift and the "+1" frame after the frameshift.

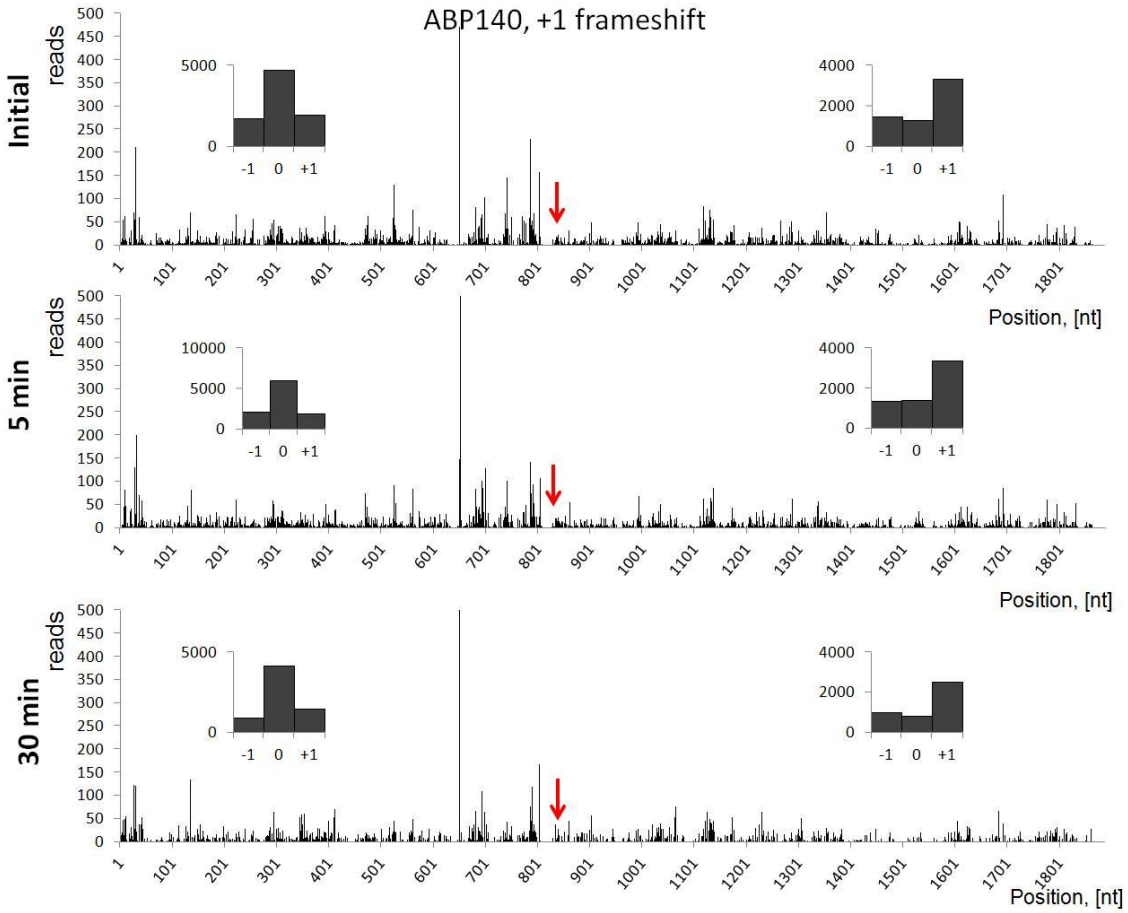

Figure 3.9. Validation of the frameshift in ABP140 gene. The 5′ ends of footprints were mapped to the genomic sequence of antizyme. The red arrow indicates the frameshift position. (Insets) Histograms show footprint counts, matching one of three possible frames either to the left or to the right of the frameshift. The "0" frame is the frame with the annotated start codon. The greatest number of footprints matched the "0" frame before the frameshift and the "+1" frame after the frameshift. Thus, we observed a change of frame, leading to the translation of a longer protein.

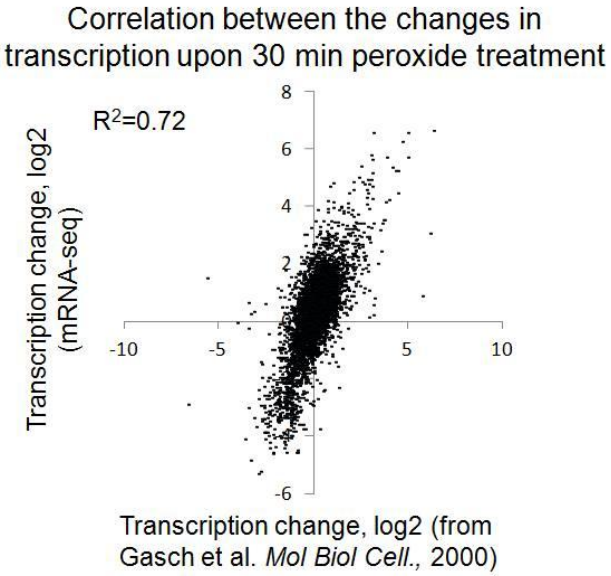

Figure 3.10. Comparison of gene expression between our RNA-seq data and the Gasch et. al. microarray data at the 30 min time point. Peroxide concentration used in our study was 0.2 mM, and in the Gasch et al. study 0.32 mM. Microarray data were taken from the online supplement of [82].

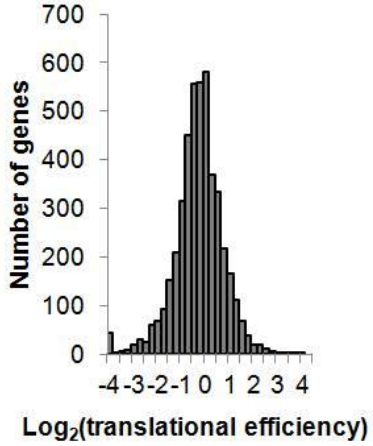

Figure 3.11. Histogram of TE shown as $\log_2$(number of footprints over the number of reads from RNA-seq).

One of our major goals was to examine genome-wide translational changes and posttranscriptional regulation of translation in response to oxidative stress. Sequencing of ribosomal footprints enabled direct and absolute quantification of mRNAs undergoing translation. It should be noted that Ribo-seq does not provide protein concentrations but instead estimates the relative translation for a given protein. Using this method, we showed that protein synthesis cannot be inferred securely from mRNA abundance. There were genes whose translation did not correlate with mRNA abundance (Fig. 3.11). In addition, a significant fraction of genes showed essentially no translation, although their mRNAs were present. We detected translational response for 97 genes after the 5-min hydrogen peroxide treatment. Only four genes showed decreased protein synthesis at this time point. After 30 min, relative protein synthesis was decreased in 593 genes and increased in 766 (Dataset S2, online supplement). Some proteins increased expression between 5 and 30 min, some reached a plateau at 5 min, and others declined during the longer treatment time.

Interestingly, the values of translation change in response to hydrogen peroxide did not match those for mRNA transcripts exactly, even if we only consider co-regulated genes (Fig. 3.12 A, black dots), although in most cases the changes in values are in the same direction. For instance, the footprint density of a representative protein increased 10-fold, but its mRNA expression increased only twofold. These data suggest a specific posttranscriptional control of protein expression. Indeed, by comparing changes in TE with changes in mRNA transcripts, we observed multiple proteins in which translational regulation was greater than transcriptional regulation (Fig. 3.12 C and D). The TE is the ratio of Riboseq read counts to mRNA-seq read counts, and it describes the propensity of

mRNA to undergo translation. The higher the TE, the better is the mRNA translated. Posttranscriptional regulation can be simply permissive, allowing an mRNA transcript to be translated under stress conditions. However, based on our analysis, posttranscriptional regulation usually makes an addition to transcription changes, modulating protein synthesis (refer to Fig. 3.1 for the TE error rate). Because we observed an immediate increase in uORF footprint density in response to hydrogen peroxide treatment, we further examined a possible effect on the TE of downstream genes. In our reference database, 3,830 genes had annotated 5′ UTRs with an unambiguous sequence longer than 23 nt. Among them, nearly 1,800 were covered by ribosomal footprints in at least one of the samples, and 1,217 had increased footprint density after the 30-in peroxide treatment. We analyzed the potential co-regulation of translation and increased ribosomal density at 5′ UTRs in these 1,800 genes and found that, on a genome-wide scale, ORF translation and ribosomal density at uORFs were mostly independent under oxidative stress conditions (Fig. 3.12 B).

Oxidative Stress Regulates Translation Elongation. We found that the density of elongating ribosomes on the mRNAs was consistently higher within the first 100–150 nt from the start codon. This observation may be explained by codon use and the corresponding tRNA copy number [89]. Hydrogen peroxide treatment caused a significant increase in ribosome occupancy and, therefore, in the density of footprint coverage within the beginning of the ORF (Fig. 3.13 A), and this effect was similar for the 5- and 30-min treatment samples. Treatment affected transcripts regardless of their length or expression level [similar to the previous observations [80]]. The data suggest that oxidative stress influenced elongation, forcing ribosomes to spend more time at the

beginning of their ORFs. Together with the increased utilization of the 5′ UTRs it explains the contradiction with previous experimental observations [76]. The fact that ribosome density increased so rapidly upon addition of hydrogen peroxide implies a direct effect of the oxidant, which targets ribosomes and elongation factors.

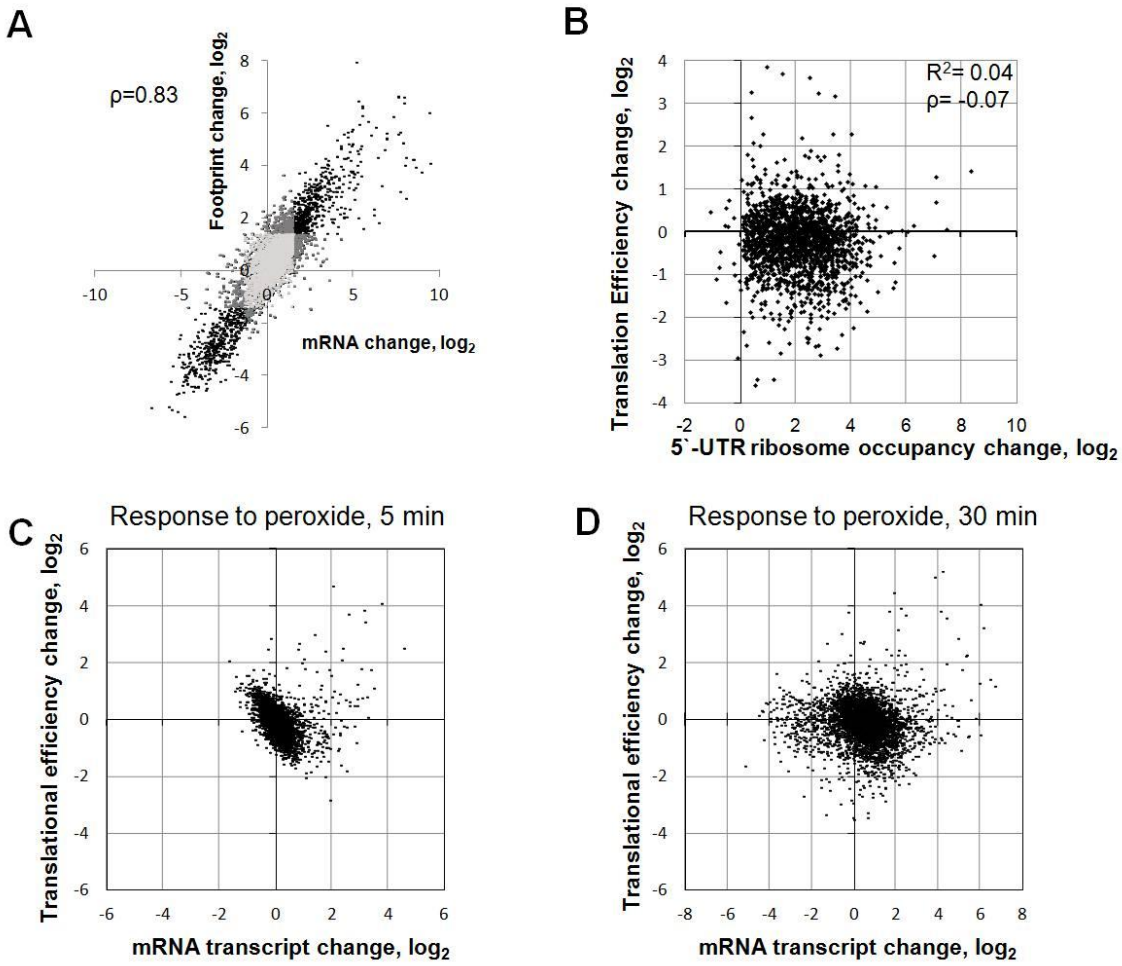

Figure 3.12. Interplay between translation and transcription. (**A**) Correlation between changes in footprint and transcript abundances in response to hydrogen peroxide. Light gray dots represent genes whose footprint count and mRNA count were not affected by peroxide treatment; dark gray dots represent genes with only the footprint or mRNA affected; and black dots represent coaffected genes. Changes in transcript and in footprint abundance between the initial and the 30-min peroxide samples are plotted on the axes (for further details see Methods). (**B**) Increased ribosomal occupancy at the 5′ UTR does not affect the TE of a downstream gene. (**C**) Relationship between change in TE and change in mRNA transcript change after 5-min incubation with peroxide. (**D**) Relationship between change in TE and change in mRNA transcript after 30-min incubation with peroxide.

Ribo-Seq Enables Codon Occupancy Quantification in Vivo. Because Ribo-seq can track translation at a single-nucleotide resolution, we examined the experimental relative frequency of translated codons and compared the experimental observations with the predicted values. Assuming that all codons are translated at the same rate, one would expect the distribution of codons trapped at the ribosomal A site to be identical to the frequency distribution of codons across mRNAs (normalized to expression levels). However, our experimental data showed that some codons were more enriched (Fig. 3.13 B, bars above the baseline), meaning that they are met more frequently in ribosomes and are translated less efficiently. Codons such as CAC or GGT fit into the relative synonymous codon use (RSCU) table, which is used for calculations of the codon adaptation index [90] that rely partially on tRNA copy numbers in the yeast genome [89]. The number of experimental replicates does not allow us to compare a particular codon directly in untreated and peroxide-treated yeast. Nevertheless, by analyzing the whole distribution (Fig. 3.13 B), we observed that the difference between predicted and experimental codon occupancy was less in stressed than in unstressed yeast. In other words, untreated, logarithmically grown yeast cells have more selective pressure on translation machinery (e.g., the availability of charged tRNA). Oxidative stress causes a rapid decrease in translation but, perhaps, less of a decrease in the pool of tRNAs and in the amount of mRNA, thus relaxing the competition of ribosomes for tRNAs. Therefore, the observed codon occupancy tends to be similar to the codon distribution of genes. Increasing the number of experimental replicates can make this method sensitive enough to detect changes in individual codon translation upon stress or any other change in condition.

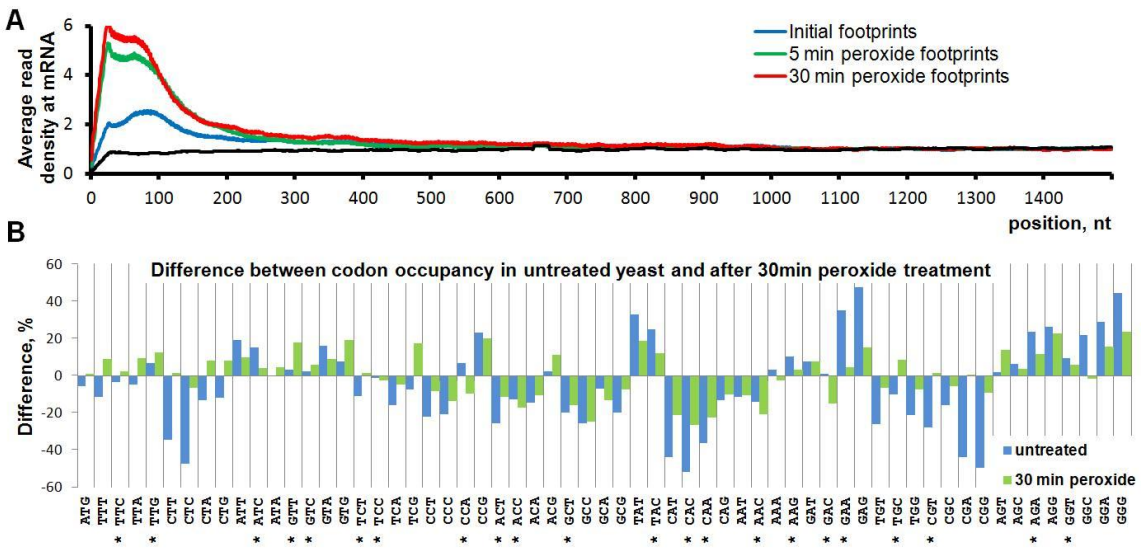

Figure 3.13. Global features of translation examined by Ribo-seq. (**A**) Density of footprint coverage along the mRNA. Profiles of read coverage were calculated for each mRNA longer than 1,500 nt and rpkm >10. The profiles were normalized based on the average density in the region from 1,000–1,500 nt. Densities for each nucleotide position were averaged across all mRNAs. An average between the two experimental replicates is shown. (**B**) Ribosomal occupancy of individual codons measured in vivo. Percentage of difference is calculated between the predicted codon distribution across mRNAs and the experimental codon appearance at the ribosomal A site. For details of normalization and prediction, see Methods. Codon values greater than zero are encountered more often at the A site of ribosome. Asterisks mark codons with the highest RSCU values [90]. Blue bars represent untreated control; green bars represent samples treated for 30 min with 0.2 mM hydrogen peroxide.

3.5 Discussion

Our data define the landscape of translational control of oxidative stress in yeast. We made several interesting observations. First, we found widespread translation of uORFs under conditions of oxidative stress. A dramatic increase in uORF ribosome occupancy occurred only 5 min after the addition of hydrogen peroxide and greatly exceeded the overall changes in protein translation. Comparisons between our study and the previously identified uORFs under conditions of starvation revealed a more extensive use of the 5′ UTRs under oxidative stress. Two times as many genes showed increased ribosome

occupancy at their 5′ UTRs under oxidative stress than under starvation (Fig. 3.14). The greater fraction of ribosomes bound to the 5′ UTR regions may be caused by two opposite events. First, translation of these regions may be up-regulated, thus producing short, cryptic peptides. On the other hand, ribosomes may move to the 5′-UTR regions slowly, accumulating footprints without affecting polypeptide yields. We think the second explanation is more likely. It agrees with the elevated density in the first 30–50 codons within the mRNA and with the reported increase in elongation time under conditions of oxidative stress [76]. It also is consistent with the complex relationships between gene translation and 5′UTR translation. Mass-spectrometry analyses would show the real yield of uORF-produced peptides and would be useful for the development of future Ribo-seq applications. We did not detect up- or down-regulation of known translation initiating factors at the 5-min time point, so the observed effects on the 5′UTR likelywere caused by posttranslational modifications of initiation factors or ribosomal proteins. Phosphorylation of eIF2, a component of the ternary initiation complex, is known to inhibit translation initiation and, consequently, protein expression [5]. In some cases, this factor was shown to induce translation of proteins, such as ATF4 or GCN4, through the intricate system of translation and reinitiation events at the uORFs [20, 91, 92]. An additional reason for increased ribosome occupancy at the 5′ UTRs may involve initiation at non-AUG codons. The majority of our detected uORFs had no AUG start codons. eIF1 and eIF5 are the factors that control the recognition of start codons during translation initiation in eukaryotes [93, 94]. We suggest that hydrogen peroxide impairs the fidelity of these factors, which normally restrict initiation to AUG codons, thereby facilitating non-AUG initiation of translation as the ribosome scans the mRNA. Our

observations imply a mechanism that slows down the ribosome at uORFs and the beginnings of ORFs. It can be achieved by impairing the exchange of elongation factors, incomplete dissociation of initiation factors, or binding additional stress-activated proteins. In addition, hydrogen peroxide may damage tRNAs [95], amino acids [96] and aminoacyl-tRNA synthetases [97]. The exact molecular mechanism requires further studies.

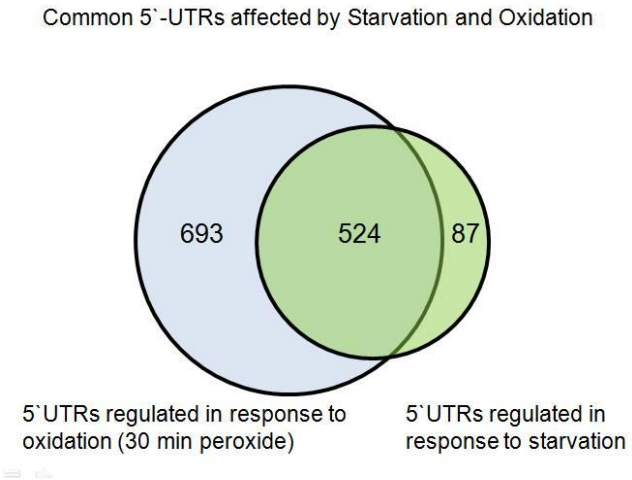

Figure 3.14. Comparison of ribosome occupancies at the 5′ UTRs affected by oxidative stress and starvation. Data for starvation were calculated by the procedure used to calculate oxidative stress. Raw sequencing files were taken from ref. [80].

Translatome and transcriptome in yeast are regulated conjointly in response to various stresses, such as amino acid depletion, osmotic shock, and sorbitol treatment [77]. Thus the genes up-regulated at the level of transcription also yield more protein product as well, a process that is termed “potentiation”. However, in the response to hydrogen peroxide only ~15% of transcriptionally regulated genes were believed to be linked by potentiation [76]. Our data indicate that the overlap is greater and that oxidative stress is not unique in this respect (Fig. 3.12 A). We compared our results directly with the published reports on the translation response to oxidative stress [76]. All proteins with high scores from that study were present in our list, and the two studies also had several down-regulated proteins in common. However, about 70% of peroxide-regulated proteins

from that study did not overlap with our hits, perhaps because the greatly increased ribosomal density at the 5′ UTRs and at the beginnings of regular ORFs, which does not reflect the actual increase of translation, compromises the microarray-based approaches. In this regard, Ribo-seq has an advantage over microarrays. Overall, our study offers a more detailed view of the translational response to oxidative stress and leads to reevaluation of many translational targets of peroxide. We also observed a significant difference between mRNA abundance and its translation (Fig. 3.11). Some mRNAs were not translated at all. Several genes had remarkably permissive posttranscriptional regulation upon hydrogen peroxide treatment. For example, Srx1, coding for sulfiredoxin, is present in unstressed yeast cells as a moderately transcribed gene with no detectable ribosomal occupancy. Its translation increases immediately after the addition of peroxide, increasing the TE by orders of magnitude. Srx1 reduces cysteinesulfinic acid, formed upon reaction with hydrogen peroxide in the active sites of peroxiredoxins. Among them, Tsa1 is one of the major proteins contributing to stress resistance [98]. An opposite example is PAB1, a poly(A)-binding protein mediating interactions between the 5′ cap structure and the 3′ mRNA poly(A) tail and facilitating translation. Treatment with hydrogen peroxide greatly decreased the TE of PAB1, but its transcript abundance remained unchanged.

Importantly, the degree of translational response to hydrogen peroxide did not match the transcriptional response precisely. There are multiple cases of posttranscriptional regulation in addition to the general transcriptional response. For example, 5-min incubation with the oxidant increased the TE of 32 genes and decreased the TE of 13. A longer incubation up-regulated 62 genes and down-regulated 122 (Dataset S1,[99]). This

finding highlights our incomplete understanding of molecular mechanisms controlling gene expression. Increasing numbers of high-throughput studies involving *S. cerevisiae* and mammalian cells that address an interplay between translation and transcription suggest that these processes do not correlate perfectly with each other in either single-cell or culture-wide conditions [73, 100, 101].

Ribo-seq offers an improved experimental alternative to the codon adaptation index [90]. It is able to detect differences between the TEs of synonymous codons. Ribo-seq may become a valuable tool for addressing the effects of deliberate starvation and amino acid depletion on codon-specific translation. Overall, our study defined the genome-wide regulation of translation by oxidative stress.

CHAPTER 4

Reassessing Eukaryotic Translation in Ribosomal Profiling Experiments

4.1 Abstract

Ribosomal profiling and high-throughput sequencing provide unprecedented opportunities for the analysis of mRNA translation. Using this novel method, several studies have demonstrated the widespread role of short upstream reading frames in translational control as well as the slower elongation at the beginning of open reading frames in response to stress. Concurrently, the relevance of translation inhibitors, such as cycloheximide, was discussed for it notably affected ribosome coverage profiles. In this study, we investigated the impact of cycloheximide over a wide concentration range on ribosome profiles in *Saccharomyces cerevisiae* and demonstrated that oxidative stress, heat shock and amino acid starvation did not affect translation elongation. We also observed little evidence for short upstream reading frames to be involved in protein synthesis regulation under stress conditions. (The conclusions made in this chapter impacts our study in Chapter 3).

4.2 Introduction

Ribosomal profiling is a common designation for several methods that examine protein translation *in vivo* by characterizing mRNA transcripts engaged in interaction with active ribosomes. A key advance in this approach was made recently by isolating mRNA fragments ("footprints") from actively translating ribosomes and subjecting them to high-throughput sequencing (Ribo-seq) [80] (Fig. 4.1).

Recent explosion of interest in the use of Ribo-seq to address numerous questions related to translation demonstrates the remarkable potential of this method. Several Ribo-seq studies reported novel and unexpected features of protein translation in yeast and mammals. For example, the ribosome distribution along mRNA was not uniform: there was a larger fraction of ribosomes residing at the beginning of transcripts, 100-200 nucleotides downstream of the start codon in yeast, pointing to slower elongation in this region. Another novel feature attributed to translational control was the widespread use and highly increased ribosomal occupancy at short upstream open reading frames (uORF) in response to amino acid starvation [80]. A study from our group showed a similar outcome under conditions of oxidative stress [99]. It was also reported that ribosomal occupancy increases immediately downstream of the start codon as a function of heat shock stress in mammalian cell cultures [102]. However, shortly after introducing Ribo-seq, some concerns have been raised regarding ribosome distribution on mRNA. It was suggested, that peaks of footprint densities is a result of cycloheximide inflicted accumulation of ribosomes, when the drug is added to cell culture [103]. In yeast, when the drug is not supplemented until the cell lysis, the peaks were significantly lower. There was no much difference in mammalian cells.

Here, we investigate how cycloheximide distorts footprint coverage across mRNA transcripts and demonstrate that the intensity of ribosome accumulation strongly depends on the intensity of stress. We found no evidence of translation elongation being affected by various stress types.

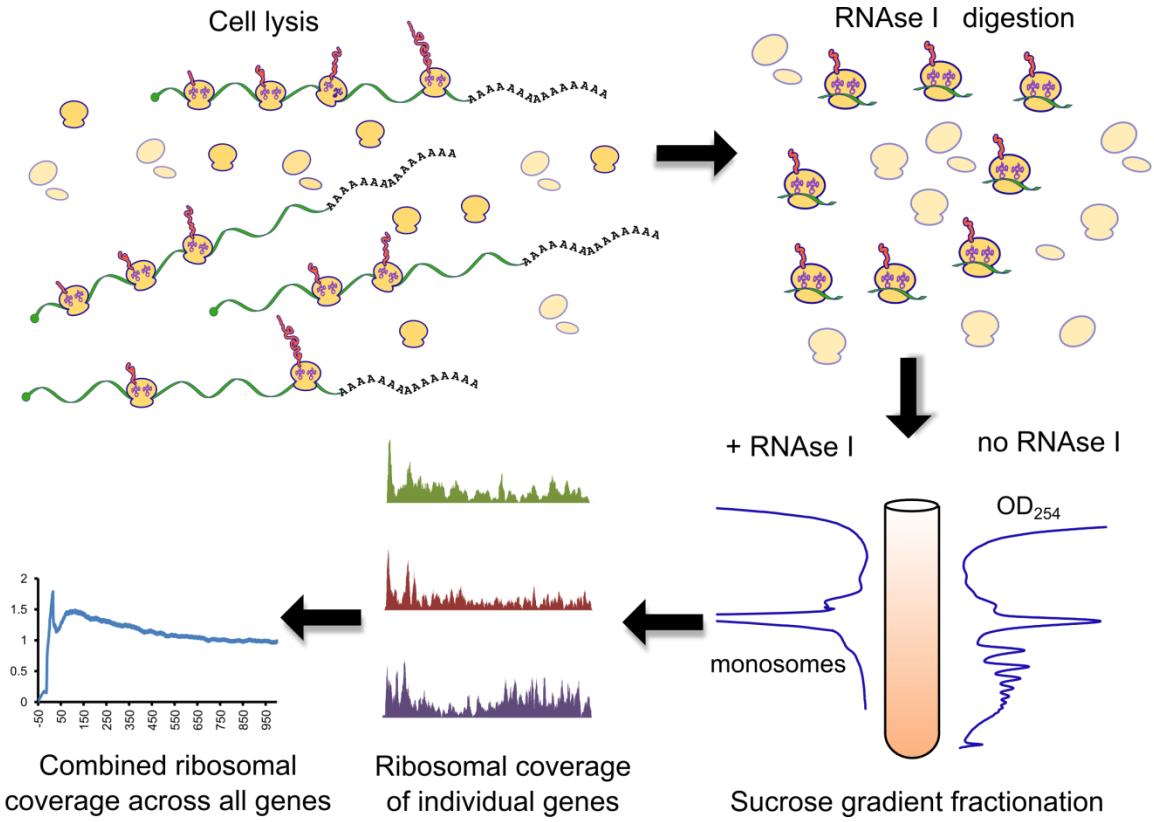

Figure 4.1. Ribosomal profiling. Cell lysis releases a mixture of individual ribosomal subunits, assembled ribosomes in complex with mRNA and blank ribosomes with no RNA attached. Sucrose gradient fractionation allows separation and isolation of these components. Captured mRNA fragments are then sequenced on an Illumina platform.

4.3 Methods

Cells and treatments. Saccharomyces cerevisiae strain BY4741 was grown on YPD agar plates for several days prior to experiments. Unless otherwise stated, the day before the experiment cells were transferred to a 50 ml flask of YPD medium and grown overnight at 30 °C with shaking. A part of that culture was inoculated into 500 ml of fresh YPD at the initial OD_{600} = 0.025 and incubated at 30 °C with shaking until the OD_{600} reached 0.5-0.6. If cultures were designated for cyloheximide treatment, the drug

was added at the end of any additional stress-inducing incubation. Immediately after drug addition, cells were harvested by vacuum filtration on 65 um filters (Millipore). It took exactly 5 min to collect the cells, which then were snap frozen in liquid nitrogen. If no drug treatment was needed, yeast cells were collected in the same manner, but filtration was initiated 5 min before the stress had to finish (see below).

Cycloheximide treatment. Concentrations of cycloheximide ranging from 1.56 to 10,000 μg/ml were used. We refer to 100 μg/ml as "x1", because it was used to inhibit protein translation in all other studies cited in this report. Therefore, other concentrations were marked as x1/64, x1/16, x1/4, x8, x100. For low concentrations, we used a stock solution in water. To achieve x8 concentration, we prepared the stock solution in DMSO. The highest possible concentration x100 was the most challenging. We first collected yeast cells by filtration, rapidly resuspended them in 5 ml of fresh YPD medium and added 5 ml of YPD with 20 mg/ml cycloheximide. This is the highest concentration possible considering drug solubility in water based solvents.

Oxidative stress. To induce oxidative stress, cells were treated with 0.2 mM hydrogen peroxide. Cell culture was incubated for 30 min at 30 °C with shaking before harvesting.

Amino acid starvation. Yeast cells were cultured in SC medium instead of YPD. This medium consists of YNB (nitrogen source), CSM (amino acids plus other supplements) and glucose as a carbon source. When the OD_{600} reached 0.5-0.6 units, cells were harvested by filtration and transferred to a new flask of pre-warmed medium containing YNB and glucose only. After 20 min incubation, cells were harvested by filtration again and snap frozen. In some cases, leucine, histidine, methionine and uracil were supplemented individually (Fig. 4.3).

Heat shock. Upon reaching the OD_{600} of 0.5-0.6 in YPD, cells were harvested by filtration and transferred to a new flask of YPD preheated to 42 °C. Incubation lasted 20 min, then cells were harvested again by the same method and frozen.

Cell lysis and ribosome isolation. Frozen cell paste pellets were cryogenically grinded by chromium beads in stainless steel vials with the aid of Mini Bead Beater (BioSpec). To achieve complete cell disruption, 10 sec pulverization cycle was applied, then vials were frozen in liquid nitrogen and an entire procedure repeated 6 more times. The content of vials was frozen during pulverization. 1 ml of ice cold lysis buffer (20 mM Tris-HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$, 1% Triton-X100, 100 μg/ml cycloheximide) was used to resuspend the pulverized cell powder. The lysates were clarified by centrifugation for 5 min. Absorbance at 260 nm was measured and 30 OD_{260} units were treated with 600 Units of RNAse I (Life Tech, Ambion) for 1 h at room temperature. The lysates were loaded on top of 10-50% sucrose gradient, prepared in the lysis buffer with no Triton. Ultracentrifugation in SW-41 Ti rotor for 3 h at 35000 rpm and 4 °C separated large ribosomal complexes from other cellular components. We used Brandel piercing system coupled with Biorad UV absorbance detector to collect monosome-containing fractions of sucrose gradients.

Footprint extraction. Sucrose fractions were concentrated with 100 kDa Amicon filter units (Millipore) to the volume of about 50-100 μl. The flow-through fraction was discarded. The volume of retentate was brought to 500 μl with release buffer (20 mM Tris-HCl pH 7.0, 6 mM EDTA, 40 U/ml Superase-In from Ambion) followed by 5 min incubation on ice. Samples were centrifuged for 5 min at 12,000 g and the flow-through

fraction was collected this time. RNA footprints were purified by a single round of acid phenol chlorophorm extraction with subsequent RNA precipitation.

Precipitation of RNA and DNA. The following components were added to the initial sample: 1/10 volume of ammonium acetate (Ambion), 5 μl of glycogen (5 mg/ml, Ambion) and 2.5 volumes of absolute ethanol. The mixture was incubated for 1 h at -20 °C, and nucleic acids were precipitated by a 15 min centrifugation. This method was used to precipitate DNA and RNA after all enzymatic reactions.

Sequencing library preparation. Footprints were treated with T4 polynucleotide kinase (Thermo Scientific) for 1 h in 10 μl total reaction volume. RNA was loaded on a 15% TBE-Urea polyacrylamide gel (Invitrogen). The band corresponding to 25-32 nt was cut out of the gel, crushed with a disposable pestle (Kimble Chase), and RNA was eluted during a 3 h incubation at 37 °C in 0.3 ml of elution buffer (20 mM Tris-HCl pH 7.0, 2 mM EDTA, 1/10 volume of 3 M ammonium acetate, 40 U/ml Superase-In). Gel particles were eliminated by a Corning Costar spin-X 0.22 μm column. RNA was precipitated. Because various libraries were prepared over extended period of times, adapter sequences, reverse transcription primers and PCR primers were different. We used two sets of primers. The first set was designed to attach 6 nucleotide barcode to the 5' end of a footprint before sequencing; another set used indexes, which were not a part of a read. Primer sequences are listed in Table 4.1 under Set #1 and Set #2.

100 ng of 3' adapter per sample were ligated by T4 RNA ligase 2 truncated KQ or K227Q (New England Biolabs). Reaction products were precipitated and reverse transcription was set up as follows. RNA pellet was dissolved in 11.5 μl of water, 4 pmol of reverse transcription primer were added along with 1 μl of dNTP mix (10 mM each).

The mixture was incubated for 5 min at 65 °C, then chilled on ice. 2 μl of DTT, 4 μl of First Strand Buffer (refer to manufacturer`s protocols for composition) were added along with 0.5 μl of SuperScript II and 0.5 μl Superase-In to the total volume of 20 μl. Reaction then continued for 30 min at 42 °C, 1 min at 65 °C, and 5 min at 80 °C. RNA was degraded in the presence of 80 mM NaOH at 95 °C for 30 min, then neutralized by the same amount of HCl. Reaction products were precipitated and run on 10% TBE-Urea polyacrylamide gel (Invitrogen), and the band corresponding to the transcriptase extended product was cut off. DNA was eluted as described above, precipitated and later used in CircLigase II reaction (20 μl total volume, according to manufacturer`s instructions, Epicentre). Products of ligation were used in the final PCR without additional precipitations or purifications. PCR was set up as follows: 10 pmol Reverse and Forward primers, 10 μl HF Buffer, 0.5 μl Phusion polymerase (New England Biolabs), 1-4 μl of ligation product, water up to 50 μl. PCR mix was subjected to 6-10 cycles of amplification (94 °C for 15 sec, 55 °C for 10 sec, 65 °C for 10 sec). Cycling was finalized for 2 min at 65 °C. PCR products were precipitated and run on 8% TBE polyacrylamide gel, and the band corresponding to the amplified products was cut off. DNA was eluted as described above, precipitated and sent for sequencing on an Illumina HiSeq 2000 platform.

Footprint alignment. Bowtie software v. 0.12.7 [84] was used to align footprints to *Saccharomyces cerevisiae* S288C genome, downloaded from SGD database with annotations. Custom Perl scripts were implemented to preprocess alignment files and plot ribosomal occupancy.

Ribosomal occupancy distribution plot. We selected all single exon genes longer than 1000 nucleotides expressed at rpkm > 30. They were aligned by start codon and coverage at every nucleotide position of every gene was averaged. The plot covers 1000 nucleotides within reading frame plus 50 nucleotides upstream of the start codon. The average coverage density of the last 300 nucleotides was used to normalize ribosome occupancy so that each profile approached the value of 1.0. We used the entire footprint sequence to calculate coverage, therefore the profile line appears smooth. Alternatively, only 5′ or 3′ ends of footprints could be used, then the profile would be more irregular.

Aligning footprints to 5′ UTR. We used a list of 5′ UTR coordinates from [87]. Additionally, for each gene we added up to 50 nt upstream of the start codon if the record in [87] annotated a shorter sequence. An extra check was done to ensure the absence of overlaps between 5′ UTR of a gene with a 3′ end of an adjacently located gene.

Table 4.1. Sequences of primers and adapters used for library preparation.

Set #1	
3' adapter	rAppCTGTAGGCACCATCAAT/3ddC/
RT-primers (barcoded)	pCGTGATGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pTGGTCAGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pATTGGCGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pACATCGGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pCACTGTGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pGCCTAAGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pTCAAGTGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pGATCTGGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pAAGCTAGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pGTAGCCGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
	pTACAAGGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG

	pCTGATCGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG
PCR forward	CAAGCAGAAGACGGCATACGA
PCR reverse	AATGATACGGCGACCACCGA
Set #2	
3' adapter	AppAGATCGGAAGAGCACACGTCT/3ddC/
RT-primer	pGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
PCR forward	AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGACGATC
PCR reverse primers (indexed)	CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**ACATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**GCCTAAG**TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**TGGTCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**CACTGT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**ATTGGC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**GATCTG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**TCAAGT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**CTGATC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**AAGCTA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**AAGCTA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	CAAGCAGAAGACGGCATACGAGAT**TACAAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

4.4 Results and Discussion

Published studies on ribosomal profiling of yeast and mammalian cells have been done with different sample preparation methods, which complicate the direct comparison. Therefore, we reproduced some treatments and stresses using budding yeast as a model organism. *Saccharomyces cerevisiae* cells were tested for amino acid starvation (as in [80]), oxidative stress (as in [99]) and heat shock (as in [102]), with and without drug treatment in the culture medium. As expected, we observed a different distribution of ribosomal occupancy, when the broad peak downstream of the start codon lowered in the absence of the drug. Unexpectedly, there was no increase in response to stress (Fig. 4.2 A-C). Therefore, none of the stress conditions tested targeted the translation elongation step.

To examine how the drug influences the ribosomal distribution, we performed a series of experiments with the cycloheximide concentrations ranging from 1.56 to 10,000 μg/ml in the medium. The shape of the occupancy peak was not constant (Fig. 4.2 F, G). The most surprising result was the disproportional decrease in ribosomal occupancy under stress. In other words, the artifactual input was not constant and was highly dependent on stress intensity.

This finding points to a passive diffusion model of cycloheximide entering live cells. The drug diffuses in a concentration-dependent manner, e.g., it takes up to 2 min to reach the equilibrium between the "in" and "out" cycloheximide concentrations [104]. The data suggest that the cycloheximide concentration does not immediately reach the threshold, under which all ribosomes are inhibited with 100% efficiency, instead increasing gradually. Therefore, following the treatment some ribosomes initiating translation

continue protein synthesis until they encounter the drug, leading to a broad cumulative peak in the ribosomal occupancy profile. The area under the peak is increased under conditions of acute stress, which leads to a steep decrease in the translation initiation rate, supposedly increasing the ratio of initiating/elongating ribosomes. Thus, the effect of drug treatment becomes more pronounced and reflects stress intensity rather than being a feature of translation (Fig. 4.2).

Based on Ribo-seq data, several studies proposed that an enrichment of "slow" codons right after the start codon is responsible for the general decrease in elongation rate [89]. A recent *in silico* study questioned this hypothesis, but still viewed the peak of ribosomal occupancy as a feature of translation in eukaryotes [105]. Our data, however, suggests departure from the slow translation model, as the increase in ribosomal occupancy at the 5' proximities of genes is dramatically less prominent than previously thought. There was only a residual slope, descending more than 300 nucleotides downstream of the start codon, implying a limited influence of a ramp of rare codons or other factors. In addition, stress does not change the ribosomal distribution (Fig. 4.2). Therefore, we think that the degree of ribosomal deceleration is not sufficient to regulate protein translation and does not play a notable role in stress response.

Mammalian cells typically show no broad peak in the absence of stress regardless of drug treatment [106]. However, heat shock increases ribosomal occupancy in the presence of the drug [102]. By analogy with yeast, this is likely an artifact, and additional profiling experiments should bring clarity on this issue. The reason the effect is more prominent in yeast is the presence of a thick cell wall serving as an additional barrier for passive diffusion. There are some obstacles with omitting drug treatments in mammalian

cell culture for they typically grow attached to a surface of a culture flask. Unlike rigid yeast cells, mammalian cells cannot be subjected to harsh mechanical interventions without losing cell integrity. Therefore, we recommend using the higher concentration of cycloheximide or any other translation inhibitor while performing ribosome profiling experiments.

Another finding of our study relates to the abundance and characteristics of uORFs. It has been shown that amino acid starvation leads to a significant increase in ribosome occupancy at uORFs [80]. A similar behavior was reported for meiosis and oxidative stress [99, 107]. However, Ribo-seq carried out without cycloheximide treatment revealed no evidence for a general increase in uORFs regardless of stress type (Fig. 4.2 D, E). A typical cycloheximide treatment provides plenty of time for free ribosomes to initiate translation and get stalled at the start codon by cycloheximide. In other words, an increase in uORF occupancy during stress co-occurs with the accumulation of footprints at the beginning of regular ORFs: it is likely artifactual and depends on stress intensity. There is no general trend supporting the use of uORFs in hundreds of genes in response to stress, although certain genes do depend on this type of regulation, such as *GCN4* transcription factor.

In conclusion, we show that slow uptake of translation inhibitor by live cells is the major cause of both ribosomal accumulation at the beginning of coding sequences and the observed widespread use of uORFs. We recommend avoiding pretreatment of cell cultures with cycloheximide or other translation inhibitors, so that post-transcriptional regulation is not perturbed. Alternatively, the concentration of inhibitors has to be as high as possible, if omitting it is not an option. The ratio of bi- and mono-valent ions in the

lysis buffer also affects results, mostly by accumulating or depleting footprints at the 5' proximal end of transcripts. Varying these factors, one can obtain different patterns of uORF and 5'-end usage, supporting different models of translation. Moreover, during stress, the artifactual ribosomal distribution changes disproportionally, reflecting stress intensity rather than adaptation at the level of protein translation.

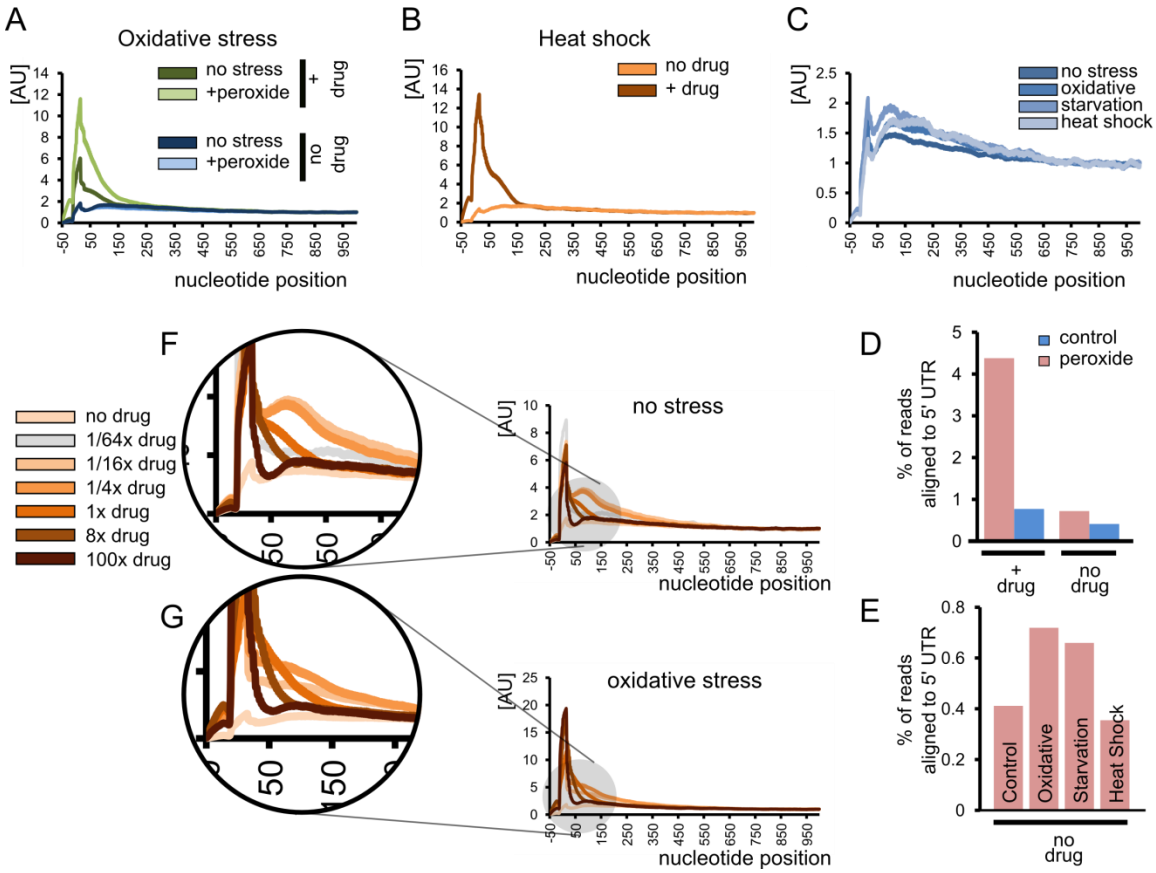

Figure 4.2. Ribosomal occupancy profiles and the effect of stress and drug treatment. (**A**) Control yeast cells and cells treated with hydrogen peroxide (0.2 mM) in the presence or absence of 100 µg/ml cycloheximide in the media. Nucleotide position count is relative to start codon. (**B**) Ribosome occupancy profiles of yeast cells undergoing heat shock (42 °C, 20 min). The peak appears only when cycloheximide is added to the medium. (**C**) None of the three tested stress types lead to a significant increase of ribosomes at the 5' proximities of reading frames in the absence of cycloheximide. Refer to Fig. 4.3 for additional details on amino acid starvation. (**D**) There is a dramatic difference in uORF occupancy if cycloheximide treatment is omitted. (**E**) None of the three examined stresses significantly increase uORF ribosomal occupancy. Although oxidative stress and amino acid starvation do slightly increase uORF occupancy, the effect is minimal compared to what was previously found [80, 99]. (**F, G**) Concentration of cycloheximide in the medium affects the shape of the profile. Cells were grown in YPD medium in the absence of stress (**F**) or subjected to oxidative stress (0.2 mM hydrogen peroxide, 30 min) (**G**).

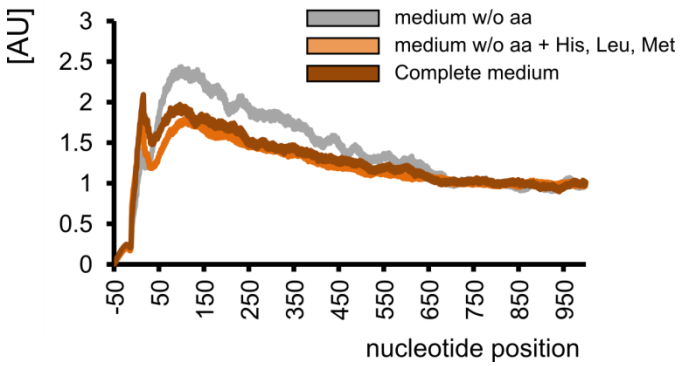

Figure 4.3.
In-depth investigation of amino acid starvation and changes in the ribosomal profile.
Repeating the experiment as was done in [80] but without cycloheximide pretreatment still leads to a slightly different ribosomal occupancy profiles (grey and dark brown lines on the graph). However, the yeast strain BY4741, used in that study is auxotroph in histidine, leucine and methionine, which are used as selective markers. Depletion of culture medium of all amino acids cannot be considered as starvation, because the lack of three essential amino acids will lead to cell death rather than to metabolism switching towards synthesis of its own amino acids. Therefore, we supplemented the medium without amino acids with normal levels of His, Met and Leu. As a result, the difference in ribosomal profiles between starved and non-starved conditions disappeared. Thus, amino acid starvation does not cause the accumulation of ribosomes at the beginning of ORFs or uORFs. The only scenario when this accumulation was observed is the absolute lack of essential amino acids, leading to ribosome stalling at the corresponding codons. This is very extreme case, which has little in common with regulation *per se*.

CHAPTER 5

FUTURE PERSPECTIVES

5.1 Ribosomal profiling in systems with extremely low net translation

Here I would like to discuss some opportunities offered by ribosome profiling, unmatched by other methods. Next generation sequencing (NGS) methods allow working with minute amounts of starting biological material. Most of the NGS applications include at least one PCR step that amplifies the template to a sufficient level. Therefore, studies in systems with extremely low translation activity are now possible. The yeast stationary phase is one of such experimental models.

Obtaining high-quality footprints from yeast in stationary phase is a challenging task. It is primarily due to the very low translational activity, wherein net translation is up to 100-fold lower than in the logarithmic growth phase [108]. Nevertheless, cells retain a substantial amount of inactive ribosomal subunits, making the footprint to rRNA ratio much lower than in the log phase. My first attempt to isolate footprints and prepare a sequencing library yielded only ~ 1% of reads that align to mRNA while the rest (99% reads) were of rRNA origin. Subtractive hybridization, similar to the one utilized in a previous study [99], helped to achieve ~10% of footprints in a library, which was not satisfactory, considering an increase in sequencing depth required to overcome high proportion of rRNA contaminants. Therefore, we took advantage of the fact that inactive eukaryotic ribosomes dissociate into subunits when exposed to a high concentration of monovalent salts [109]. By elevating KCl concentration in sucrose gradients during ultracentrifugation to 0.5 M, we were able to isolate only translationally active

ribosomes, omitting individual large and small subunits and preventing them from reassociation upon cell lysis (Fig. 5.1).

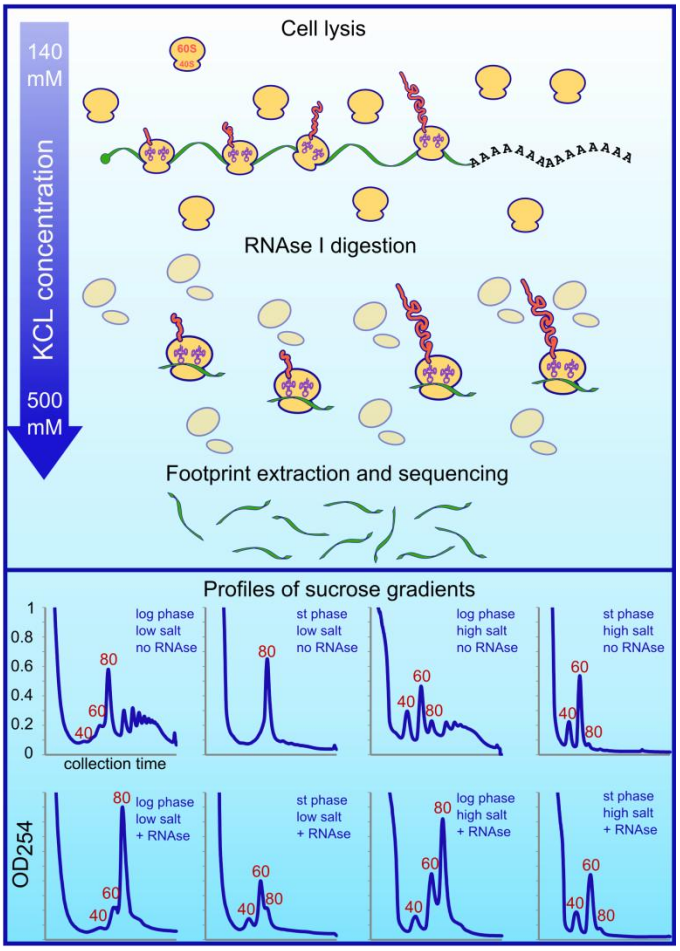

Figure 5.1 Method outline. The upper panel shows isolation of ribosomal footprints. The lower panel illustrates the influence of high salt (500 mM KCl) and RNAse digestion on sucrose gradient profiles. Peaks corresponding to small and large subunits and the assembled ribosome are marked as 40, 60 and 80 S. The bottom rightmost panel displays conditions used to isolate footprints in this study. By dissociating blank, inactive ribosomes into small and large subunits, we collected only the 80S fraction containing active ribosomes. Thus, substantial rRNA contamination was avoided.

Samples from stationary yeast cultures treated this way displayed the footprint yield comparable to that obtained from log phase culture. Subsequent experiments demonstrated that 0.3 M KCl is sufficient to dissociate yeast ribosomes (Fig. 5.2 A). This manner of footprint isolation has both strengths and weaknesses. It notably depletes ribosomes from the 5′ ends of mRNA transcripts. Thus, we observed a decline in ribosomal occupancy immediately after the start codon (Fig. 5.2 A). This also makes the search for short upstream ORF (uOFR) irrelevant. On the other hand, it does not affect gene expression estimate if the first 100 nt after the start codon are omitted during

calculations (r^2=0.94). Protein reading frames also show sharper codon periodicity pattern in footprint coverage (Fig. 5.2B).

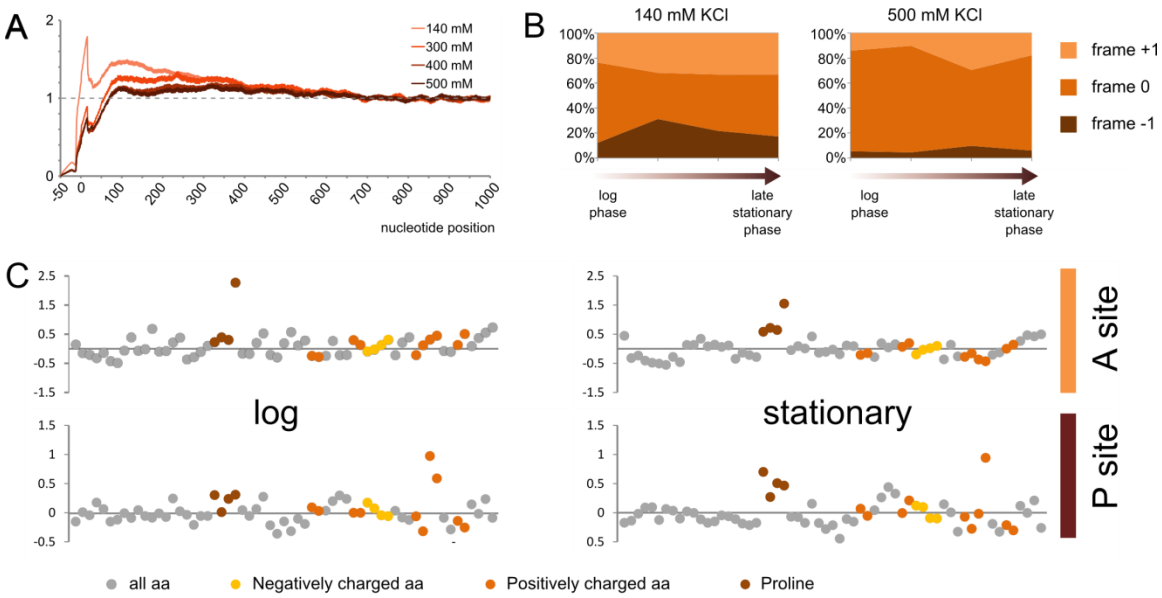

Figure 5.2 Footprint coverage in log and stationary growth phase. (**A**) Influence of increasing KCl concentrations on the ribosomal density profile. Profiles of read coverage were calculated for each mRNA longer than 1000 nt and rpkm > 30. The profiles were normalized based on the average density in the 700-1000 nt region. Densities for each nucleotide position were averaged across all mRNAs. An average between 2 replicates is shown for each salt concentration. (**B**) High concentration of KCl in a sucrose gradient buffer results in a distinct pattern of codon periodicity. Ribosomal profiling allows tracking protein reading frames by looking at the fraction of footprints supporting that frame. Because ribosome moves 3 nucleotides at a time, footprint coverage of mRNA also shows a clear periodic pattern. Typically, we map the 3` ends of footprins on mRNA and count how many of them fit into each of the three possible frames. Stacked area charts show the percentage of reads fitted into the three reading frames for samples prepared with low and high salt concentrations in sucrose gradient buffer, which is used during ultracentrifugation. Only verified genes with no introns were used to generate the charts. Thus, the majority of footprints has to support the reading frame 0. As can be seen, lower ionic strength sometimes yields footprints without clear periodic pattern, equally distributed across all reading frames. This requires additional bioinformatic steps to determine the actual frame, such as selection of footprints of certain length. On the other hand, high ionic strength results in sharp profiles in all samples. These results were reproducible across various experimental conditions. (**C**) Scatter plots of ribosomal occupancy at individual codons inferred from ribosomal A and P sites. Two time points are displayed: log phase and the latest stationary phase. 61 sense codons are plotted (grey dots). Values above the zero baseline correspond to the codons with higher chance of being occupied by ribosome. Zero value would be assigned to a codon whose chances of being occupied by ribosome are the same as its occurrence in the transcriptome. Codons

corresponding to negatively and positively charged amino acids are colored in yellow and orange. Proline codons are shown in brown.

Based on footprint`s average length of 28 nucleotides and strict codon periodicity pattern, I projected locations of A (acceptor) and P (peptide) sites of the ribosome active core. Since this experimental condition has dozens of millions reads collected, a large portion of the yeast transcriptome is covered at high depth. I estimated the relative frequencies of each codon in A or P ribosomal sites and compared them to the frequencies of codons across known open reading frames. If the ribosome has no preference for a particular codon, both frequencies would be equal. However, if the ribosome pauses or stalls at this triplet, an increased number of overlaying footprints is expected. I found "slow" codons in the A site, which require more time to translate (Fig. 5.2C). The group of proline codons were more represented in the stationary phase in the P site of ribosome. Besides this observation, there was no striking difference between log and stationary phase yeast cells. This finding suggests that even under severe nutrient deficiency none of the individual codons serve as a rate-limiting bottleneck for protein translation in general. As a positive control of such extreme ribosomal stalling, I performed a histidine depletion assay, wherein the same yeast strain (histidine auxotroph) was grown in complete SD medium during the log phase and then transferred to the medium with no histidine. In this case, ribosomes stalled at CAU and CAC histidine codons (Fig 5.3). These experiments illustrate the remarkable potential of ribosome profiling and set future directions for me to pursue.

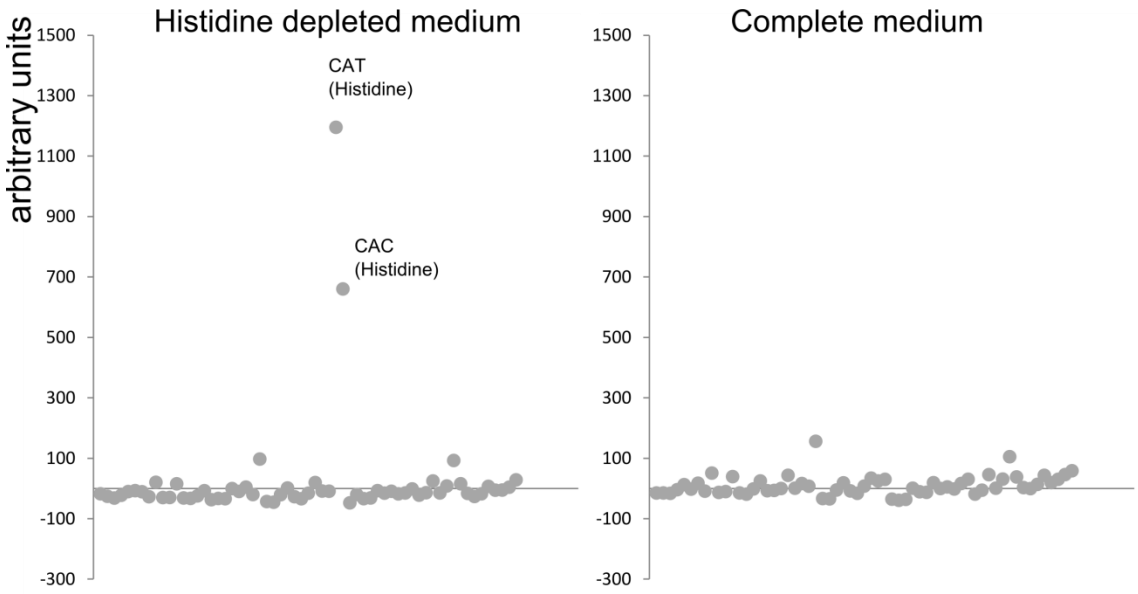

Figure 5.3 Scatter plots of ribosomal occupancy at individual codons inferred from A sites of a ribosome. Each dot corresponds to a codon. Yeast were grown in a complete medium and then transferred to a medium without histidine. Ribosome stalled at histidine codons, because the yeast strain used in this experiment is deficient in histidine synthesis.

5.2 Specialized ribosome hypothesis

As I mentioned in the introduction, ribosome is a large complex consisting of many proteins. Some of them are not crucial for performing the main function – synthesizing polypeptide chains. However, they theoretically can modify the fidelity of translation, recognition of start and stop codons on the mRNA, restrict certain mRNAs from getting into the active center of ribosome, etc. Studies have revealed that ribosomal complexes from various tissues have different molar ratios of their protein subunits. Thus, ribosomes are not identical across the tissues and potentially may control protein translation even when faced with similar mRNA pools. This control is referred to as the "ribosomal code" [1, 110]. Some progress has been made in this area, which revealed that certain mRNA species are differentially expressed due to ribosome composition [110, 111]. The number

of disease phenotypes attributed to ribosomal abnormalities is growing. For instance, mutations in ribosomal protein L38 (RPL38) were recently identified as responsible for tissue-specific pattering defects in mouse development [112]. In zebrafish, mutations in several RPs are associated with distinct phenotypic defects in brain, eye, ear and body [113]. In most of these cases the mechanisms that would account for tissue specific phenotypes remain poorly understood or have been attributed to extra ribosomal protein functions [113]. A special interest is paid to human disease associated mutations in ribosomal proteins. In particular, Diamond-Blackfan anemia is the most striking example for which the direct link between the disease phenotype and the ribosomal dysfunction was established. Personalized genome sequencing of thousands patients has recently provided evidence of co-occurrence of ribosomal point mutations with various cancer types [114, 115]. Nevertheless, a lack of *in vivo* methods prevented testing the ribosomal code theory on a genome-wide scale. The emerging methods of ribosome profiling perfectly fit this niche and studies on functional diversity of ribosomes would thrive in the nearest future.

1. Xue, S. and M. Barna, *Specialized ribosomes: a new frontier in gene regulation and organismal biology.* Nat Rev Mol Cell Biol, 2012. **13**(6): p. 355-69.
2. Ben-Shem, A., et al., *The structure of the eukaryotic ribosome at 3.0 A resolution.* Science, 2011. **334**(6062): p. 1524-9.
3. Fleischer, T.C., et al., *Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes.* Genes Dev, 2006. **20**(10): p. 1294-307.
4. Dever, T.E. and R. Green, *The elongation, termination, and recycling phases of translation in eukaryotes.* Cold Spring Harb Perspect Biol, 2012. **4**(7): p. a013706.
5. Jackson, R.J., C.U. Hellen, and T.V. Pestova, *The mechanism of eukaryotic translation initiation and principles of its regulation.* Nat Rev Mol Cell Biol, 2010. **11**(2): p. 113-27.
6. Lu, J. and A. Holmgren, *Selenoproteins.* J Biol Chem, 2009. **284**(2): p. 723-7.
7. Stock, T. and M. Rother, *Selenoproteins in Archaea and Gram-positive bacteria.* Biochim Biophys Acta, 2009. **1790**(11): p. 1520-32.
8. Lobanov, A.V., D.L. Hatfield, and V.N. Gladyshev, *Eukaryotic selenoproteins and selenoproteomes.* Biochim Biophys Acta, 2009. **1790**(11): p. 1424-8.
9. Wilbur, B., ed. *Molecular biology of the gene*. 6 ed. 2008, Pearson Education Inc.: San Francisco.
10. Hinnebusch, A.G., *Molecular mechanism of scanning and start codon selection in eukaryotes.* Microbiol Mol Biol Rev, 2011. **75**(3): p. 434-67, first page of table of contents.
11. Kozak, M., *Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes.* Cell, 1986. **44**(2): p. 283-92.
12. Kozak, M., *An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs.* Nucleic Acids Res, 1987. **15**(20): p. 8125-48.
13. Kozak, M., *At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells.* J Mol Biol, 1987. **196**(4): p. 947-50.
14. Peabody, D.S., *Translation initiation at non-AUG triplets in mammalian cells.* J Biol Chem, 1989. **264**(9): p. 5031-5.
15. Kolitz, S.E., J.E. Takacs, and J.R. Lorsch, *Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation.* RNA, 2009. **15**(1): p. 138-52.
16. Clements, J.M., T.M. Laz, and F. Sherman, *Efficiency of translation initiation by non-AUG codons in Saccharomyces cerevisiae.* Mol Cell Biol, 1988. **8**(10): p. 4533-6.
17. Starck, S.R., et al., *Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I.* Science, 2012. **336**(6089): p. 1719-23.
18. Chang, K.J. and C.C. Wang, *Translation initiation from a naturally occurring non-AUG codon in Saccharomyces cerevisiae.* J Biol Chem, 2004. **279**(14): p. 13778-85.
19. Tang, H.L., et al., *Translation of a yeast mitochondrial tRNA synthetase initiated at redundant non-AUG codons.* J Biol Chem, 2004. **279**(48): p. 49656-63.
20. Hinnebusch, A.G., *Translational regulation of GCN4 and the general amino acid control of yeast.* Annu Rev Microbiol, 2005. **59**: p. 407-50.

21. Zhang, F. and A.G. Hinnebusch, *An upstream ORF with non-AUG start codon is translated in vivo but dispensable for translational control of GCN4 mRNA.* Nucleic Acids Res, 2011. **39**(8): p. 3128-40.
22. Thoreen, C.C., et al., *A unifying model for mTORC1-mediated regulation of mRNA translation.* Nature, 2012. **485**(7396): p. 109-13.
23. Fabian, M.R., N. Sonenberg, and W. Filipowicz, *Regulation of mRNA translation and stability by microRNAs.* Annu Rev Biochem, 2010. **79**: p. 351-79.
24. Arnâer, E.S., *Focus on mammalian thioredoxin reductases--important selenoproteins with versatile functions.* Biochimica et Biophysica Acta: Protein Structure and Molecular Enzymology, 2009. **1790**(6): p. 495-526.
25. Holmgren, A. and M. Bjèornstedt, *Thioredoxin and thioredoxin reductase.* Methods in Enzymology, 1995. **252**: p. 199-208.
26. Gladyshev, V.N., K.T. Jeang, and T.C. Stadtman, *Selenocysteine, identified as the penultimate C-terminal residue in human T-cell thioredoxin reductase, corresponds to TGA in the human placental gene.* Proc. Natl. Acad. Sci. U.S.A., 1996. **93**(12): p. 6146-51.
27. Bjèornstedt, M., S. Kumar, and A. Holmgren, *Selenite and selenodiglutathione: reactions with thioredoxin systems.* Methods in Enzymology, 1995. **252**: p. 209-19.
28. Zhong, L. and A. Holmgren, *Essential role of selenium in the catalytic activities of mammalian thioredoxin reductase revealed by characterization of recombinant enzymes with selenocysteine mutations.* Journal of Biological Chemistry, 2000. **275**(24): p. 18121-8.
29. May, J.M., et al., *Reduction of dehydroascorbate to ascorbate by the selenoenzyme thioredoxin reductase.* Journal of Biological Chemistry, 1997. **272**(36): p. 22607-10.
30. Andersson, M., A. Holmgren, and G. Spyrou, *NK-lysin, a disulfide-containing effector peptide of T-lymphocytes, is reduced and inactivated by human thioredoxin reductase. Implication for a protective mechanism against NK-lysin cytotoxicity.* Journal of Biological Chemistry, 1996. **271**(17): p. 10116-20.
31. Jakupoglu, C., et al., *Cytoplasmic thioredoxin reductase is essential for embryogenesis but dispensable for cardiac development.* Molecular and Cellular Biology, 2005. **25**(5): p. 1980-8.
32. Conrad, M., et al., *Essential role for mitochondrial thioredoxin reductase in hematopoiesis, heart development, and heart function.* Molecular and Cellular Biology, 2004. **24**(21): p. 9414-23.
33. Sun, Q.A., et al., *Redox regulation of cell signaling by selenocysteine in mammalian thioredoxin reductases.* Journal of Biological Chemistry, 1999. **274**(35): p. 24522-30.
34. Sun, Q.A., et al., *Reaction mechanism and regulation of mammalian thioredoxin/glutathione reductase.* Biochemistry, 2005. **44**(44): p. 14528-37.
35. Su, D., et al., *Mammalian selenoprotein thioredoxin-glutathione reductase. Roles in disulfide bond formation and sperm maturation.* Journal of Biological Chemistry, 2005. **280**(28): p. 26491-8.
36. Sun, Q.A., et al., *Selenoprotein oxidoreductase with specificity for thioredoxin and glutathione systems.* Proc. Natl. Acad. Sci. U.S.A., 2001. **98**(7): p. 3673-8.

37. Kehr, S., et al., *X-ray fluorescence microscopy reveals the role of selenium in spermatogenesis.* Journal of Molecular Biology, 2009. **389**(5): p. 808-18.
38. Ursini, F., et al., *Dual function of the selenoprotein PHGPx during sperm maturation.* Science, 1999. **285**(5432): p. 1393-6.
39. Fernandes, A.P. and A. Holmgren, *Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system.* Antioxid. Redox. Signal., 2004. **6**(1): p. 63-74.
40. Shelton, M.D., P.B. Chock, and J.J. Mieyal, *Glutaredoxin: role in reversible protein s-glutathionylation and regulation of redox signal transduction and protein translocation.* Antioxid. Redox. Signal., 2005. **7**(3-4): p. 348-66.
41. Holmgren, A., *Antioxidant function of thioredoxin and glutaredoxin systems.* Antioxid. Redox. Signal., 2000. **2**(4): p. 811-20.
42. Rundlèof, A.K., et al., *Evidence for intriguingly complex transcription of human thioredoxin reductase 1.* Free Radic. Biol. Med., 2004. **36**(5): p. 641-56.
43. Su, D. and V.N. Gladyshev, *Alternative splicing involving the thioredoxin reductase module in mammals: a glutaredoxin-containing thioredoxin reductase 1.* Biochemistry, 2004. **43**(38): p. 12177-88.
44. Kanzok, S.M., et al., *Substitution of the thioredoxin system for glutathione reductase in Drosophila melanogaster.* Science, 2001. **291**(5504): p. 643-6.
45. Cheng, Z., et al., *The relationship of the redox potentials of thioredoxin and thioredoxin reductase from Drosophila melanogaster to the enzymatic mechanism: reduced thioredoxin is the reductant of glutathione in Drosophila.* Biochemistry, 2007. **46**(26): p. 7875-85.
46. Alger, H.M. and D.L. Williams, *The disulfide redox system of Schistosoma mansoni and the importance of a multifunctional enzyme, thioredoxin glutathione reductase.* Molecular and Biochemical Parasitology, 2002. **121**(1): p. 129-39.
47. Angelucci, F., et al., *Glutathione reductase and thioredoxin reductase at the crossroad: the structure of Schistosoma mansoni thioredoxin glutathione reductase.* Proteins, 2008. **72**(3): p. 936-45.
48. Bonilla, M., et al., *Platyhelminth mitochondrial and cytosolic redox homeostasis is controlled by a single thioredoxin glutathione reductase and dependent on selenium and glutathione.* Journal of Biological Chemistry, 2008. **283**(26): p. 17898-907.
49. Peabody, D.S., *Translation initiation at non-AUG triplets in mammalian cells.* J. Biol. Chem., 1989. **264**(9): p. 5031-5.
50. Wegrzyn, J.L., et al., *Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation initiation sites.* BMC bioinformatics, 2008. **9**: p. 232.
51. Kieft, J.S., *Viral IRES RNA structures and ribosome interactions.* Trends in Biochemical Sciences, 2008. **33**(6): p. 274-83.
52. Baird, S.D., et al., *Searching for IRES.* RNA, 2006. **12**(10): p. 1755-85.
53. Sonenberg, N. and A.G. Hinnebusch, *Regulation of translation initiation in eukaryotes: mechanisms and biological targets.* Cell, 2009. **136**(4): p. 731-45.
54. Kozak, M., *Initiation of translation in prokaryotes and eukaryotes.* Gene, 1999. **234**(2): p. 187-208.

55. Arnâer, E.S., et al., *High-level expression in Escherichia coli of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the selA, selB and selC genes.* Journal of Molecular Biology, 1999. **292**(5): p. 1003-16.
56. Mokrejs, M., et al., *IRESite: the database of experimentally verified IRES structures (www.iresite.org).* Nucleic Acids Research, 2006. **34**(Database issue): p. D125-30.
57. Turanov, A.A., D. Su, and V.N. Gladyshev, *Characterization of alternative cytosolic forms and cellular targets of mouse mitochondrial thioredoxin reductase.* J Biol Chem, 2006. **281**(32): p. 22953-63.
58. Rundlof, A.K., M. Carlsten, and E.S. Arner, *The core promoter of human thioredoxin reductase 1: cloning, transcriptional activity, and Oct-1, Sp1, and Sp3 binding reveal a housekeeping-type promoter for the AU-rich element-regulated gene.* Journal of Biological Chemistry, 2001. **276**(32): p. 30542-51.
59. Rundlof, A.K., et al., *Quantification of alternative mRNA species and identification of thioredoxin reductase 1 isoforms in human tumor cells.* Differentiation, 2007. **75**(2): p. 123-32.
60. Chang, E.Y., et al., *Induction of apoptosis by the overexpression of an alternative splicing variant of mitochondrial thioredoxin reductase.* Free Radic. Biol. Med., 2005. **39**(12): p. 1666-75.
61. Miranda-Vizuete, A. and G. Spyrou, *Genomic organization and identification of a novel alternative splicing variant of mouse mitochondrial thioredoxin reductase (TrxR2) gene.* Mol. Cells, 2002. **13**(3): p. 488-92.
62. Dammeyer, P., et al., *Induction of cell membrane protrusions by the N-terminal glutaredoxin domain of a rare splice variant of human thioredoxin reductase 1.* Journal of Biological Chemistry, 2008. **283**(5): p. 2814-21.
63. Damdimopoulou, P.E., et al., *The human thioredoxin reductase-1 splice variant TXNRD1_v3 is an atypical inducer of cytoplasmic filaments and cell membrane filopodia.* Biochimica et Biophysica Acta: Protein Structure and Molecular Enzymology, 2009.
64. Arnaud, E., et al., *A new 34-kilodalton isoform of human fibroblast growth factor 2 is cap dependently synthesized by using a non-AUG start codon and behaves as a survival factor.* Mol Cell Biol, 1999. **19**(1): p. 505-14.
65. Bonnal, S., et al., *Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site trans-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA.* J Biol Chem, 2005. **280**(6): p. 4144-53.
66. Starck, S.R., et al., *A distinct translation initiation mechanism generates cryptic peptides for immune surveillance.* PLoS One, 2008. **3**(10): p. e3460.
67. Xiao, J.H., et al., *Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1.* Cell, 1991. **65**(4): p. 551-68.
68. Falvey, E., F. Fleury-Olela, and U. Schibler, *The rat hepatic leukemia factor (HLF) gene encodes two transcriptional activators with distinct circadian rhythms, tissue distributions and target preferences.* EMBO Journal, 1995. **14**(17): p. 4307-17.

69. Imataka, H., H.S. Olsen, and N. Sonenberg, *A new translational regulator with homology to eukaryotic translation initiation factor 4G.* EMBO Journal, 1997. **16**(4): p. 817-25.
70. Taira, M., et al., *A human testis-specific mRNA for phosphoribosylpyrophosphate synthetase that initiates from a non-AUG codon.* Journal of Biological Chemistry, 1990. **265**(27): p. 16491-7.
71. Hinkson, I.V. and J.E. Elias, *The dynamic state of protein turnover: It's about time.* Trends Cell Biol, 2011. **21**(5): p. 293-303.
72. Huh, W.K., et al., *Global analysis of protein localization in budding yeast.* Nature, 2003. **425**(6959): p. 686-91.
73. Newman, J.R., et al., *Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise.* Nature, 2006. **441**(7095): p. 840-6.
74. Bar-Even, A., et al., *Noise in protein expression scales with natural protein abundance.* Nat Genet, 2006. **38**(6): p. 636-43.
75. Arava, Y., et al., *Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3889-94.
76. Shenton, D., et al., *Global translational responses to oxidative stress impact upon multiple levels of protein synthesis.* J Biol Chem, 2006. **281**(39): p. 29011-21.
77. Halbeisen, R.E. and A.P. Gerber, *Stress-Dependent Coordination of Transcriptome and Translatome in Yeast.* PLoS Biol, 2009. **7**(5): p. e105.
78. Arava, Y., et al., *Dissecting eukaryotic translation and its control by ribosome density mapping.* Nucleic Acids Res, 2005. **33**(8): p. 2421-32.
79. Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities.* Nat Rev Genet, 2011. **12**(2): p. 87-98.
80. Ingolia, N.T., et al., *Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.* Science, 2009. **324**(5924): p. 218-23.
81. Ingolia, N.T., *Genome-wide translational profiling by ribosome footprinting.* Methods Enzymol, 2010. **470**: p. 119-42.
82. Gasch, A.P., et al., *Genomic expression programs in the response of yeast cells to environmental changes.* Mol Biol Cell, 2000. **11**(12): p. 4241-57.
83. Fourcroy, P., et al., *Polyribosome analysis on sucrose gradients produced by the freeze-thaw method.* J Biochem Biophys Methods, 1981. **4**(3-4): p. 243-6.
84. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
85. Warner, J.R., *The economics of ribosome biosynthesis in yeast.* Trends Biochem Sci, 1999. **24**(11): p. 437-40.
86. Lawless, C., et al., *Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast.* BMC Genomics, 2009. **10**: p. 7.
87. Nagalakshmi, U., et al., *The transcriptional landscape of the yeast genome defined by RNA sequencing.* Science, 2008. **320**(5881): p. 1344-9.
88. Asakura, T., et al., *Isolation and characterization of a novel actin filament-binding protein from Saccharomyces cerevisiae.* Oncogene, 1998. **16**(1): p. 121-30.
89. Tuller, T., et al., *An evolutionarily conserved mechanism for controlling the efficiency of protein translation.* Cell, 2010. **141**(2): p. 344-54.

90. Sharp, P.M. and W.H. Li, *The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.* Nucleic Acids Res, 1987. **15**(3): p. 1281-95.
91. Vattem, K.M. and R.C. Wek, *Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells.* Proc Natl Acad Sci U S A, 2004. **101**(31): p. 11269-74.
92. Lu, P.D., H.P. Harding, and D. Ron, *Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response.* J Cell Biol, 2004. **167**(1): p. 27-33.
93. Nanda, J.S., et al., *eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation.* J Mol Biol, 2009. **394**(2): p. 268-85.
94. Ivanov, I.P., et al., *Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1).* Proc Natl Acad Sci U S A, 2010. **107**(42): p. 18056-60.
95. Thompson, D.M., et al., *tRNA cleavage is a conserved response to oxidative stress in eukaryotes.* RNA, 2008. **14**(10): p. 2095-103.
96. Stadtman, E.R. and R.L. Levine, *Free radical-mediated oxidation of free amino acids and amino acid residues in proteins.* Amino Acids, 2003. **25**(3-4): p. 207-18.
97. Ling, J. and D. Soll, *Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site.* Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4028-33.
98. Biteau, B., J. Labarre, and M.B. Toledano, *ATP-dependent reduction of cysteine-sulphinic acid by S. cerevisiae sulphiredoxin.* Nature, 2003. **425**(6961): p. 980-4.
99. Gerashchenko, M.V., A.V. Lobanov, and V.N. Gladyshev, *Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress.* Proc Natl Acad Sci U S A, 2012. **109**(43): p. 17394-9.
100. Lu, R., et al., *Systems-level dynamic analyses of fate change in murine embryonic stem cells.* Nature, 2009. **462**(7271): p. 358-62.
101. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.
102. Shalgi, R., et al., *Widespread regulation of translation by elongation pausing in heat shock.* Mol Cell, 2013. **49**(3): p. 439-52.
103. Ingolia, N.T., et al., *The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments.* Nat Protoc, 2012. **7**(8): p. 1534-50.
104. Shearer, G., Jr. and P.S. Sypherd, *Cycloheximide efflux in antibiotic-adapted cells of the fungus Mucor racemosus.* Antimicrob Agents Chemother, 1988. **32**(3): p. 341-5.
105. Shah, P., et al., *Rate-limiting steps in yeast protein translation.* Cell, 2013. **153**(7): p. 1589-601.
106. Ingolia, N.T., L.F. Lareau, and J.S. Weissman, *Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes.* Cell, 2011. **147**(4): p. 789-802.
107. Brar, G.A., et al., *High-resolution view of the yeast meiotic program revealed by ribosome profiling.* Science, 2012. **335**(6068): p. 552-7.

108. Fuge, E.K., E.L. Braun, and M. Werner-Washburne, *Protein synthesis in long-term stationary-phase cultures of Saccharomyces cerevisiae.* J Bacteriol, 1994. **176**(18): p. 5802-13.
109. Zylber, E.A. and S. Penman, *The effect of high ionic strength on monomers, polyribosomes, and puromycin-treated polyribosomes.* Biochim Biophys Acta, 1970. **204**(1): p. 221-9.
110. Gilbert, W.V., *Functional specialization of ribosomes?* Trends Biochem Sci, 2011. **36**(3): p. 127-32.
111. Lee, A.S., R. Burdeinick-Kerr, and S.P. Whelan, *A ribosome-specialized translation initiation pathway is required for cap-dependent translation of vesicular stomatitis virus mRNAs.* Proc Natl Acad Sci U S A, 2013. **110**(1): p. 324-9.
112. Kondrashov, N., et al., *Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning.* Cell, 2011. **145**(3): p. 383-97.
113. Uechi, T., et al., *Ribosomal protein gene knockdown causes developmental defects in zebrafish.* PLoS One, 2006. **1**: p. e37.
114. Shenoy, N., et al., *Alterations in the ribosomal machinery in cancer and hematologic disorders.* J Hematol Oncol, 2012. **5**: p. 32.
115. Luft, F., *The rise of a ribosomopathy and increased cancer risk.* J Mol Med (Berl), 2010. **88**(1): p. 1-3.